The Dissertation Committee for Xinlei Wang
certifies that this is the approved version of the following dissertation:

Bayesian Variable Selection for GLM

Committee:

Thomas Shively, Supervisor

Edward George, Supervisor

Betsy Greenberg

William Jefferys

Thomas Sager

S.Lynne Stokes

Bayesian Variable Selection for GLM

by

Xinlei Wang, B.S., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2002

Dedicated to my family.

Acknowledgments

I would like to thank my advisors, Professor Edward George, and Professor Thomas Shively for their advice and support throughout my graduate career.

I would also like to thank the other members of my committee - Professor Betsy Greenberg, Professor William Jefferys, Professor Thomas Sager and Professor S.Lynne Stokes - for their valuable discussions and suggestions.

I would like to thank my husband, Min Chen, for his help in R programming and LaTeXusage. I would like to thank my parents and my parents-in-law for their generous family support and encouraging words.

Finally, I would like to thank professors at Department of Statistics, University of Pittsburgh, who initiated my strong interests in Statistics. Thanks to Dr. Leon Gleser, who welcomed me into the Bayesian world.

Bayesian Variable Selection for GLM

Publication No. ____________________

Xinlei Wang, Ph.D.
The University of Texas at Austin, 2002

Supervisors: Thomas Shively
Edward George

I consider the problem of variable selection for Generalized Linear Models (GLM). A great deal of effort has been expended in variable selection for linear regression models and many selection criteria have been proposed and well known in practice. However, for GLM, the standard practice is to use criteria AIC or BIC, or use Chi-square tests for nested models. Due to great difficulty in achieving analytical tractability, much less research in variable selection has been done for GLM, even if it is parallel to linear regression models. In this dissertation, I present a comprehensive Bayesian solution to this problem, which extends the hierarchical formulation of George and Foster (2000) to GLM. It involves choosing priors for parameters and models that bring in hyperparameters, integrating model-specific parameters out of the likelihood function, estimating the values of the hyperparameters from data or choosing hyperpriors for the hyperparameters and finally obtaining posterior probabilities of models as selection criteria. Unlike most previous research in

this field, the model posterior achieved in this work can be calculated easily and accurately without resorting to simulation methods like the Gibbs sampling, Reversible Jump MCMC, etc., hence bypassing the high-dimensional convergence problem.

I achieve analytical tractability for GLM by proposing an Integrated Laplace Approximation that has been shown better than classical Laplace's method in this context. I describe two approaches for developing selection criteria: Empirical Bayes (EB), and Fully Bayes (FB), which are different in the way of handling hyperparameters. I also present an alternative FB approach, Conditional Fully Bayes (CFB), based on a different hyper-parameterization. In addition, I propose a method of restricting the integration region of the hyperparameters to improve FB selection performance. For each approach, various criteria are developed and their performance is evaluated through simulation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Variable Selection Problem for Generalized Linear Models

The problem of variable selection is typically stated as follows: given a dependent variable Y and a set of potential explanatory variables or predictors $X_1, X_2, \ldots, X_p$ that are thought to contain redundant or irrelevant variables, one wants to identify a subset of $X_1, X_2, \ldots, X_p$ that best models the underlying relationship revealed by the data.

Variable selection is a fundamental part of building a regression model. Indeed, under the normal linear regression context, variable selection has been the focus of considerable attention in the statistics literature (e.g., Mallows, 1973; Akaike, 1974; Hocking, 1976; Thompson, 1978; Schwarz, 1978; Draper and Smith, 1981; Weisberg, 1985; Mitchell and Beauchamp, 1988; Miller, 1990; George and McCulloch, 1993 and 1997; George and Foster, 1994 and 2000) and a wide variety of selection criteria have been proposed, which include adjusted R^2, C_p, $PRESS_p$, AIC, BIC, RIC, etc. Specifically, for normal linear regression models, the objective of variable selection is to select and fit the "best" model of the form

$$Y = X_1^*\beta_1^* + X_2^*\beta_2^* + \ldots + X_q^*\beta_q^* + \epsilon$$

where $X_1^*, X_2^*, \ldots, X_q^*$ is a "selected" subset of $X_1, X_2, \ldots, X_p$ and $\epsilon \sim N(0, \sigma^2 I)$. The reasons that the fundamental developments of variable selection occurred in the context of the linear regression model are: (1) historically, the linear regression model is most widely used in practice; many applications with non-linear relationship or non-normal data can be transformed to the setting; (2) a lot of problems of interest can be posed as linear variable selection problems (3) the linear regression model has analytic tractability that greatly facilitates insights.

However, the focus on the linear models in previous literatures does not indicate that it is redundant to consider variable selection for Generalized Linear Models (GLM). GLM has a lot of applications in disciplines as widely varied as agriculture, ecology, economics, education, medicine, psychology and sociology, etc, for which the linear regression model is not proper. Examples are applications with discrete or categorical dependent variables or the distribution of dependent variables cannot be transformed to normal ones. Hence, the variable selection problem for GLM plays an important role in practice. As people have recognized this, some recent Bayesian research emerged for GLM (e.g., Raftery and Richardson, 1993; Raftery, 1996; Clyde and DeSimone-Sasinowska, 1998; Dellaportas and Forster, 1999; Clyde, 1999; Dellaportas, Forster and Ntzoufras, 2000 and 2002; Ntzoufras, Dellaportas and Forster, 2001). Explicitly, the variable selection problem for GLM is set up as follows. Suppose that the components of $\mathbf{Y} = (y_1, y_2, \ldots, y_n)^T$ are independent random variables, in which each y_i follows an exponential family distribution as below:

$$f(y_i|\theta_i, \phi) = \exp\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\} \tag{1.1}$$

and may depend on p independent known variables $X_i, i = 1, 2, ..., p$. Letting γ index the subsets of $X_1, X_2, \ldots, X_p$ and q_γ be the size of the γth subset, the objective is to select and fit a model of the form

$$g(E(\mathbf{Y})) = \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma$$

where g is a link function, $\mathbf{X}_\gamma$ is the $n \times (q_\gamma + 1)$ design matrix whose columns correspond to the γth subset, $\boldsymbol{\beta}_\gamma$ is a $(q_\gamma+1)\times 1$ vector of regression coefficients.

The most often used variable selection criteria for GLM are based on calibration of normed likelihoods or deviances by degree of freedom that can provide a measure of the distance of each model from data. For example, the calibrated deviance can be written as $D_c = D - 2q_\gamma \log(\alpha)$. Often, $\alpha = 1/e$ that yields $D + 2q_\gamma$ and $\alpha = 1/\sqrt{n}$ that yields $D + 2q_\gamma \log n$ are reasonable choices in practice that are actually AIC and BIC.

1.2 Basic Review of GLM

This section gives readers a basic review of knowledge in GLM, which includes exponential family distributions, link functions and the likelihood function in a matrix formula under GLM with the canonical link.

1.2.1 Components of GLM

A GLM is determined by the distribution of its response variable, linear predictors and link function. Usually, the response variable y_i, $i = 1, 2, \ldots, n$ in GLM are assumed to be independent observations from an exponential family of the form defined in equation (1.1). The mean and variance of y_i are given by

$$E(y_i) = \mu_i = b'(\theta_i)$$

$$Var(y_i) = \phi \, b''(\theta_i)$$

The covariates $X_1, X_2, \ldots, X_p$ are represented by the $n \times (p+1)$ design matrix $\mathbf{X}$ and are incorporated into the model through the linear predictor $\boldsymbol{\eta} \equiv \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ and related to the means $\boldsymbol{\mu}$ via the link function g, where $\eta_i = g(\mu_i)$, $i = 1, 2, \ldots, n$.

As mentioned above, GLM is restricted to members of exponential family. Two of the most common discrete distributions, Poisson and Binomial, and two common continuous distributions, Normal and Gamma are all included in this family. Table (1.1) lists characteristics of some exponential family distributions.

Link functions describe the relationship between the mean of the ith observation and its linear predictors. It can often be used to advantage to linearize seemingly nonlinear structures. For example, logistic and Gomperz growth curves become linear when respectively the logit and complementary log log links are used. In this dissertation, for simplicity but without loss of

Table 1.1: Characteristics of Some Exponential Families

Model		ϕ	θ_i	μ_i	$b(\theta_i)$	$c(y_i, \phi)$
Binomial	$B(n, p_i)$	1	$\log(\frac{p_i}{1-p_i})$	$\frac{ne^{\theta_i}}{1+e^{\theta_i}}$	$n\log(1+e^{\theta_i})$	$\log\binom{n}{y_i}$
Poisson	$P(\lambda_i)$	1	$\log(\lambda_i)$	e^{θ_i}	e^{θ_i}	$-\log(y_i!)$
Negative Binomial	$NB(v, p_i)$	1	$\log(1-p_i)$	$\frac{ve^{\theta_i}}{1-e^{\theta_i}}$	$v\log(1-e^{\theta_i})^{-1}$	$\log\binom{v+y_i-1}{v-1}$
Normal	$N(\mu, \sigma^2)$	σ^2	μ_i	θ_i	$\theta_i^2/2$	$-\frac{y_i^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$
Gamma	$GA(\mu_i, v)$	$\frac{1}{v}$	$-\frac{1}{\mu_i}$	$-\frac{1}{\theta_i}$	$-\log(-\theta_i)$	$(\frac{1}{\phi} - 1)\log y_i - \frac{\log \phi}{\phi} - \log\Gamma(\phi^{-1})$
Inverse Gaussian	$GI(\mu_i, \sigma^2)$	σ^2	$-(2\mu_i^2)^{-1}$	$\sqrt{\frac{-1}{2\theta_i}}$	$-(-2\theta_i)^{1/2}$	$\frac{1}{2y_i\phi} - \frac{1}{2}\log(2\pi y_i^3\phi)$

generality, I focus on the GLM with the canonical link, that is $\boldsymbol{\eta} \equiv \boldsymbol{\theta} = b'^{-1}(\boldsymbol{\mu})$, where we have $g(.) = b'^{-1}(.)$. Note that in this case $Var(y_i) = \phi\ b''(\eta_i) = \phi\ b''(b'^{-1}(\mu_i))$. In Chapter 6, I show that all the work can be generalized to GLM with a noncanonical link. Without explicitly stating it, I always refer to GLM with the canonical link. Table (1.2) gives some examples of canonical link functions.

1.2.2 The Likelihood Function for GLM

Considering variable selection, suppose that only an unknown subset of the β_j coefficients are nonzero. Let $\gamma = 1, 2, \ldots, 2^p$ index all the subsets of $\{X_1, X_2, \ldots, X_p\}$, and q_γ be the number of elements in subset γ. For the

Table 1.2: Canonical Link Functions for Some Exponential Families

Distribution	**Canonical link function**	
Poisson	Log	$\eta_i = \log(\mu_i)$
Binomial	Logit	$\eta_i = \log(\frac{\pi_i}{1-\pi_i}) = \log(\frac{\mu_i}{n_i-\mu_i})$
Normal	Identity	$\eta_i = \mu_i$
Gamma	Reciprocal	$\eta_i = \frac{1}{\mu_i}$
Inverse Gaussian	Reciprocal^2	$\eta_i = \frac{1}{\mu_i^2}$

setting of GLM with the canonical link, denote

$$\mathbf{X}_\gamma = \begin{pmatrix} 1 & x_{(1),1} & x_{(2),1} & \cdots & x_{(q_\gamma),1} \\ 1 & x_{(1),2} & x_{(2),2} & \cdots & x_{(q_\gamma),2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{(1),n} & x_{(2),n} & \cdots & x_{(q_\gamma),n} \end{pmatrix}_{n\times(q_\gamma+1)}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n\times 1} \quad \boldsymbol{\beta}_\gamma = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{q_\gamma} \end{pmatrix}_{(q_\gamma+1)\times 1}$$

$$\mathbf{b}(\boldsymbol{\theta}) = \mathbf{b}(\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma) = \begin{pmatrix} b(\theta_1) \\ b(\theta_2) \\ \vdots \\ b(\theta_n) \end{pmatrix} = \begin{pmatrix} b(\ (1,\ x_{(1),1}\ ,\ x_{(2),1}\ ,\cdots\ ,\ x_{(q_\gamma),1})\cdot\boldsymbol{\beta}_\gamma) \\ b(\ (1,\ x_{(1),2}\ ,\ x_{(2),2}\ ,\cdots\ ,\ x_{(q_\gamma),2})\cdot\boldsymbol{\beta}_\gamma) \\ \vdots \\ b(\ (1,\ x_{(1),n}\ ,\ x_{(2),n}\ ,\cdots\ ,\ x_{(q_\gamma),n})\cdot\boldsymbol{\beta}_\gamma) \end{pmatrix}_{n\times 1}$$

$$\mathbf{c}(\mathbf{Y},\phi) = \begin{pmatrix} c(y_1,\phi) \\ c(y_2,\phi) \\ \vdots \\ c(y_n,\phi) \end{pmatrix}$$

where $(x_{(1)}, x_{(2)}, \cdots, x_{(q_\gamma)})$ is the subset indexed γ, and $\boldsymbol{\beta}_\gamma$ are the corresponding coefficients of model γ. Let $\mathbf{b}'(\boldsymbol{\theta})$ be the derivative of $\mathbf{b}(\boldsymbol{\theta})$ and $\mathbf{b}^T(\boldsymbol{\theta})$ be the transpose of $\mathbf{b}(\boldsymbol{\theta})$. We have

$$p(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) = \exp\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi)\cdot\mathbf{1}\} \tag{1.2}$$

Equation (1.2) is used throughout this dissertation.

1.3 A Hierarchical Bayesian Approach

There are several reasons to consider a Bayesian approach. First, as argued by Leamer (1978), traditional model selection using non-experimental data involves a great deal of personal judgment. Bayesian methods make assumptions explicit and allow for a more formal mathematical incorporation of personal opinions. Second, as will be shown below, the Bayesian framework is straightforward to implement. Third, a Bayesian approach can easily account for model uncertainty. As stated by Weisberg (1985), "There is no final model, only a group of possible models that are all judged nearly equally useful", which clearly indicates the existence of model uncertainty. A Bayesian approach can further average over a set of models through the posterior probabilities and hence can incorporate the uncertainty in subsequent inferences.

There is a hierarchical Bayesian solution to the problem of variable selection: we build into the model directly a vector of indicator variables γ that reflects which covariates are included in the model. If one assigns a prior distribution $\pi(\gamma|\boldsymbol{\psi}_1)$ to the set of possible models, where $\boldsymbol{\psi}_\mathbf{1}$ is the hyperparameter vector introduced by the prior of γ, then Bayesian updating of the prior distribution leads to a posterior distribution given the data $\mathbf{Y} =$

$(y_1, y_2, \ldots, y_n)^T$ over the different models:

$$\pi(\gamma|\mathbf{Y}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) = \frac{p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)\pi(\gamma|\boldsymbol{\psi}_1)}{\sum_{\gamma'} p(\mathbf{Y}|\gamma', \boldsymbol{\psi_2})\pi(\gamma'|\boldsymbol{\psi_1})} \tag{1.3}$$

where

$$p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2) = \int p(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)p(\boldsymbol{\beta}_\gamma|\gamma, \boldsymbol{\psi}_2)\, d\boldsymbol{\beta}_\gamma \tag{1.4}$$

is the marginal distribution of the data $\mathbf{Y}$ after integrating out model specific parameters with respect to the prior distribution $p(\boldsymbol{\beta}_\gamma|\gamma, \boldsymbol{\psi}_2)$ on $\boldsymbol{\beta}_\gamma$. Note $\boldsymbol{\psi}_2$ is the hyperparameter vector introduced by the prior of $\boldsymbol{\beta}_\gamma$ and is independent of γ.

For the hyperparameter vectors $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, there exist two alternative approaches: (1) an Empirical Bayes (EB) approach that estimates $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ based on data, and then uses $\pi(\gamma|\mathbf{Y}, \hat{\boldsymbol{\psi}}_1, \hat{\boldsymbol{\psi}}_2)$ as the variable selection criterion; (2) a Fully Bayes (FB) approach that puts priors on $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, then integrates them out and uses $\pi(\gamma|\mathbf{Y})$ as the variable selection criterion, that is,

$$\begin{aligned}
\pi(\gamma|\mathbf{Y}) &= \iint_D \pi(\gamma|\mathbf{Y}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2)p(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2|\mathbf{Y})\, d\boldsymbol{\psi}_1\, d\boldsymbol{\psi}_2 \\
&= \iint_D \frac{p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)\pi(\gamma|\boldsymbol{\psi}_1)}{p(\mathbf{Y}|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)} \cdot \frac{p(\mathbf{Y}|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)}{p(\mathbf{Y})}\, d\boldsymbol{\psi}_1\, d\boldsymbol{\psi}_2 \\
&= \iint_D \frac{p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)\pi(\gamma|\boldsymbol{\psi}_1)}{p(\mathbf{Y})} \cdot \pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\, d\boldsymbol{\psi}_1\, d\boldsymbol{\psi}_2 \qquad (1.5)
\end{aligned}$$

where $p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)$ is given by (1.4); and D is the area that contains all possible values for $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ based on the prior distribution $\pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ on $\boldsymbol{\psi}_1$,$\boldsymbol{\psi}_2$. It is reasonable to assume that the prior on $\boldsymbol{\psi}_1$ and the prior on $\boldsymbol{\psi}_2$ are independent, so we actually have $\pi(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2) = \pi(\boldsymbol{\psi}_1)\pi(\boldsymbol{\psi}_2)$.

In practice, implementation of this hierarchical Bayesian variable selection has some difficulties:

1. Specifying prior distributions requires considerable effort especially for the FB approach, in which case we need to find proper prior distributions for not only model specific parameters but also hyperparameters introduced by the prior distributions of parameters.

2. The integration required to obtain $p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)$ in (1.4) is analytically intractable for most cases of GLM. In previous research, people got around this by using approximate methods of integration such as Laplace methods or Monte Carlo methods (Kass and Raftery, 1995) or by applying normal theory results after transforming the data (Clyde, 1999).

In this dissertation, workable solutions to the above two problems are presented in the context of GLM: proper priors for $\boldsymbol{\beta}_\gamma$, $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$, and an integrated Laplace approximation to $p(\mathbf{Y}|\gamma, \boldsymbol{\psi}_2)$ are proposed, which make it feasible to compute the integrals in equation (1.3) and (1.5) accurately when the sample size is reasonably large.

1.4 Overview

In this dissertation, I investigate a Bayesian approach to the variable selection problem for GLM. The study is motivated by George and Foster (2000), which has shown for the normal linear setting, selection criteria such as AIC/C_p, BIC and RIC, which have fixed dimensionality penalties, correspond to selection of maximum posterior models under implicit hyperparameter choices for a particular hierarchical Bayesian formulation. They also proposed Empirical Bayes criteria, which have dimensionality penalties that depend on the data. Intuitively, we can conjecture that all these could be generalized to GLM under the Bayesian approach. In this dissertation, I confirm this conjecture by choosing proper priors for parameters and proper values for hyperparameters under an approximation to the posterior of models that asymptotically converges to the true posterior. Also, I propose both EB and FB approaches to achieve the posterior probabilities of models and develop various variable selection criteria. Last, I run simulations to evaluate the performance of proposed criteria and compare them with traditional criteria for GLM such as AIC and BIC.

The structure of this dissertation is as follows. Chapter 1 defines the variable selection problem for GLM, introduces the notations and provides the theoretical background. In Chapter 2, I describe the Bayesian framework for variable selection under GLM. I discuss the methods used to select prior distributions. I propose an Integrated Laplace approximation to achieve the marginal distribution of $\mathbf{Y}$ in equation (1.4) and compare it with classical Laplace approximation. Also, I present the posterior distribution given hyperparameters and calibrate it to selection criteria AIC and BIC. In Chapter 3, the EB approach to deal with hyperparameters for variable selection is presented. I generalize the CML criterion to GLM that was first proposed by

George and Foster (2000) for the normal linear regression and also propose $MCML$ and $HCML$, two different versions of CML, to enhance the performance of the EB criteria without sacrificing computational simplicity. Chapter 4 describes the FB approach based on different hyperpriors including noninformative and conjugate prior distributions. I discuss the problems with the FB selection criteria and then propose a method that restricts the integration area of hyperparameters to solve problems and improve performance. In Chapter 5, I describe an alternative FB approach called Conditional FB (CFB) that relaxes hyperparameters to model-specific ones, and present criteria based on it. In Chapter 6, I generalize all the results achieved in previous chapters to GLM with noncanonical link functions. I also discuss a special case of GLM, the normal linear regression and how the results achieved here connect to those in George and Foster (2000). Chapter 7 examines the performance of selection criteria from the EB, FB and CFB approaches through simulation. I run simulation based on Poisson, Bernoulli and Normal distribution for different sample size and different number of potential variables. Conclusions, discussions and future work are presented in Chapter 8. The calculation details of the marginal distribution of data $\mathbf{Y}$ are given in Appendix A. The calculation details of the posterior distribution of models based on restricted integration area are presented in Appendix B. The complete simulation results by number of nonzero components are listed in Appendix C.

Chapter 2

Bayesian Framework for GLM

In this chapter I describe the selection of prior distributions and propose an Integrated Laplace approximation to obtain model posterior give hyperparameters. I show, by choosing proper values of hyperparameters, that the model posterior can be calibrated to AIC and BIC. I also describe methods about how to specify the prior mean vector of model-specific parameters and the dispersion parameter ϕ.

2.1 Selection of Prior Distributions

Suppose ϕ is known, an important step in the Bayesian approach to variable selection for GLM is to choose priors for the unknown quantities γ and $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$.

In variable selection problems, prior model information often takes the form of prior evidence for the inclusion of a variable rather than an individual model, that is

$$\pi(\gamma|\omega_1, \omega_2, \ldots, \omega_p) = \prod_{i=1}^{p}[\omega_i^{\gamma_i}(1-\omega_i)^{1-\gamma_i}]$$

where ω_i, $i = 1, 2, \ldots, p$ is the prior probability that X_i is included in a model and γ_i is the indicator of whether X_i is included. Without strong prior information, I follow George and Foster (2000) by further taking $\omega_1 = \omega_2 = \ldots = \omega_p \equiv \omega$ so that each variable has equal probability of inclusion. Hence, the prior model probability is specified as

$$\pi(\gamma|\omega) = \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} \text{ for } \omega \in (0,1) \tag{2.1}$$

The remainder of this section is devoted to the consideration of the prior distribution for the model specific parameters $\boldsymbol{\beta_\gamma}$. When available, informative prior distribution for $\boldsymbol{\beta_\gamma}$ should be elicited and incorporated into the analysis. In the absence of expert opinion and previous knowledge, one might seek to choose prior distributions which reflect uncertainty about the parameters and also embody reasonable priori constraints. I adopt the following prior distribution on $\boldsymbol{\beta_\gamma}$:

$$\boldsymbol{\beta_\gamma}|\gamma, c \sim \mathbf{N}_{q_\gamma+1}(\mathbf{m}_\gamma, c\,\phi\,(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)^{-1}) \text{ for } c \in (0, +\infty) \tag{2.2}$$

where

$$\mathbf{V}_\gamma = \begin{pmatrix} \mathbf{b}''(\hat{\theta}_{\gamma 1}) & 0 & 0 & 0 \\ 0 & \mathbf{b}''(\hat{\theta}_{\gamma 2}) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{b}''(\hat{\theta}_{\gamma n}) \end{pmatrix}_{n\times n} \tag{2.3}$$

$\hat{\boldsymbol{\theta}}_\gamma = \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma$ and $\hat{\boldsymbol{\beta}}_\gamma$ is the maximum likelihood estimator of $\boldsymbol{\beta_\gamma}$ under model γ. The prior covariance matrix $U = c\,\phi\,(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)^{-1}$ corresponds to the estimated unit information matrix (see Kass and Wasserman 1995). For a generalized linear model, the estimated information matrix $\mathbf{I}(\hat{\boldsymbol{\beta}}_\gamma)$ is given by

$$\mathbf{I}(\hat{\boldsymbol{\beta}}_\gamma) = \frac{\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma}{\phi}$$

The unit information matrix is given by $\mathbf{I}(\boldsymbol{\beta}_\gamma)/N$ where N is the number of units in the data. This is typically the number of observations of $\mathbf{Y}$. Another advantage of form (2.2) is its analytical tractability for the normal case or asymptotically for other cases in the exponential family under an Integrated Laplace approximation to $p(\mathbf{Y}|\gamma, c)$, as will be shown in next section.

2.2 Approximation to the Marginal Distribution of Y

As discussed in Section 1.3, one difficulty in Bayesian variable selection for GLM is to evaluate $p(\mathbf{Y}|\gamma, c)$. Under the prior of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ given in form (2.2), we have

$$
\begin{aligned}
p\left(\mathbf{Y}|\gamma, c\right) &= \int_{\mathbf{R}^{q_\gamma+1}} p\left(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma\right) p\left(\boldsymbol{\beta}_\gamma|\gamma, c\right) d\boldsymbol{\beta}_\gamma \\
&= (2\pi)^{-\frac{q_\gamma+1}{2}} \left|\frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{c\phi}\right|^{\frac{1}{2}} \int_{\mathbf{R}^{q_\gamma+1}} \exp\left\{\frac{\mathbf{Y}^T \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma \boldsymbol{\beta}_\gamma)\cdot \mathbf{1}}{\phi}\right. \\
&\quad \left. + \mathbf{c}^T(\mathbf{Y}, \phi)\cdot \mathbf{1} - \frac{(\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma (\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)}{2c\phi}\right\} d\boldsymbol{\beta}_\gamma \qquad (2.4)
\end{aligned}
$$

Except for the normal case when the prior (2.2) is conjugate, the above integration has no closed-form solution. Hence, analytical or numerical approximation methods are needed. In section 2.2.1, I apply the classical Laplace method that plays a central role in approximating posterior probabilities and moments in previous research (Tierney and Kadane, 1986; Kass and Wasserman, 1995; Raftery, 1996); in section 2.2.2, I propose an approximation that takes advantage of the normality of the prior on $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$; and the accuracy of this method is evaluated. In section 2.2.3, these two methods are compared in a brief discussion.

2.2.1 Laplace Approximation

Assuming h_n is a smooth real function with a d-dimensional parameter $\boldsymbol{\omega}$, having a maximum at $\hat{\boldsymbol{\omega}}$, Laplace's method approximates an integral of the form

$$I = \int f(\boldsymbol{\omega}) \exp(h_n(\boldsymbol{\omega})) d\boldsymbol{\omega}$$

by expanding f and h_n around $\hat{\boldsymbol{\omega}}$. The subscript n is used here to denote that the function h_n depends, apart from $\boldsymbol{\omega}$, on a variable n,which can tend to infinity; in statistical application n is typically the sample size. The factor $\exp(h_n(\boldsymbol{\omega}))$ in the integrand is approximated by a function proportional to a normal density determined by the second-order Taylor series approximation to h_n. The Laplace approximation to I is then:

$$I_L = f(\hat{\boldsymbol{\omega}})(2\pi)^{d/2}(\det \Sigma)^{1/2} \exp\{h_n(\hat{\boldsymbol{\omega}})\} \tag{2.5}$$

where $\Sigma = -(\mathbf{D}^2\, h_n(\hat{\boldsymbol{\omega}}))^{-1}$ (minus the inverse of the Hessian of h_n evaluated $\hat{\boldsymbol{\omega}}$). The errors of the approximation are of order $O(n^{-1})$ (see Bleistein and Handelsman 1975), that is

$$I_L = I\{1 + O(n^{-1})\}$$

Applying the Laplace approximation (2.5) to (2.4) by replacing $\hat{\boldsymbol{\omega}}$ with $\hat{\boldsymbol{\beta}}_\gamma$, which is the maximum likelihood estimator of $\boldsymbol{\beta}_\gamma$ for model γ, $\exp\{h_n(\hat{\boldsymbol{\omega}})\}$

with $p\,(\mathbf{Y}|\hat{\boldsymbol{\beta}}_\gamma,\gamma)$ and $f(\hat{\boldsymbol{\omega}})$ with $p\,(\hat{\boldsymbol{\beta}}_\gamma|\gamma,c)$, we have

$$p_L\,(\mathbf{Y}|\gamma,c) = p\,(\hat{\boldsymbol{\beta}}_\gamma|\gamma,c)(2\pi)^{\frac{q_\gamma+1}{2}}\left|-\mathbf{D}^2\log p\,(\mathbf{Y}|\hat{\boldsymbol{\beta}}_\gamma,\gamma)\right|^{-1/2}\cdot p\,(\mathbf{Y}|\hat{\boldsymbol{\beta}}_\gamma,\gamma)$$

where

$$-\mathbf{D}^2\log p\,(\mathbf{Y}|\hat{\boldsymbol{\beta}}_\gamma,\gamma) = \frac{\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma}{\phi}$$

Hence,

$$\begin{aligned}
p_L(\mathbf{Y}|\gamma,c) = (2\pi)^{-\frac{q_\gamma+1}{2}}&\left|\frac{\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma}{c\phi}\right|^{\frac{1}{2}}\exp\left\{-\frac{(\hat{\boldsymbol{\beta}}_\gamma-\mathbf{m}_\gamma)^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma(\hat{\boldsymbol{\beta}}_\gamma-\mathbf{m}_\gamma)}{2c\phi}\right\}\\
&\cdot(2\pi)^{\frac{q_\gamma+1}{2}}\left|\frac{\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma}{\phi}\right|^{-\frac{1}{2}}\exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-\mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi}+\mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}\\
= \quad & c^{-\frac{q_\gamma+1}{2}}\exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-\mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi}+\mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}\\
&\exp\left\{-\frac{(\hat{\boldsymbol{\beta}}_\gamma-\mathbf{m}_\gamma)^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma(\hat{\boldsymbol{\beta}}_\gamma-\mathbf{m}_\gamma)}{2c\phi}\right\} \qquad (2.6)
\end{aligned}$$

2.2.2 An Integrated Laplace Approximation

Since I adopt the prior of $\boldsymbol{\beta}_\gamma$ as a multivariate normal distribution, it is of interest to try an approximation that would give the exact answer for the normal case. Like Laplace's method, I apply the second order approximation to the log-likelihood function by expanding it about $\hat{\boldsymbol{\beta}}_\gamma$. However, I integrate $p\,(\boldsymbol{\beta}_\gamma|\gamma,c)$ out along with the expansion of $p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma)$ rather than simply approximating it by $p\,(\hat{\boldsymbol{\beta}}_\gamma|\gamma,c)$ and then factoring it out of the integration. This method is called an Integrated Laplace approximation. Notice that

$$\left.\frac{\partial\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma)}{\partial\boldsymbol{\beta}_\gamma}\right|_{\boldsymbol{\beta}_\gamma=\hat{\boldsymbol{\beta}}_\gamma} = 0$$

We have,

$$\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma) = \frac{\mathbf{Y}^T\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\boldsymbol{\beta}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}$$

$$\approx \frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1} - \frac{1}{2}(\boldsymbol{\beta}_\gamma - \hat{\boldsymbol{\beta}}_\gamma)^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma(\boldsymbol{\beta}_\gamma - \hat{\boldsymbol{\beta}}_\gamma)}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1} \tag{2.7}$$

where $\mathbf{V}_\gamma$ is defined in (2.3). Apply approximation (2.7) to equation (2.4),

$$p\,(\mathbf{Y}|\gamma,c) \approx \tilde{p}\,(\mathbf{Y}|\gamma,c) = (2\pi)^{-\frac{q_\gamma+1}{2}}\left|\frac{\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma}{c\phi}\right|^{\frac{1}{2}} \int_{\mathbf{R}^{q_\gamma+1}} \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1} - \frac{1}{2}(\boldsymbol{\beta}_\gamma - \hat{\boldsymbol{\beta}}_\gamma)^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma(\boldsymbol{\beta}_\gamma - \hat{\boldsymbol{\beta}}_\gamma)}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1} - \frac{(\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)}{2c\phi}\right\} d\boldsymbol{\beta}_\gamma \tag{2.8}$$

After some calculations (details are given in Appendix A), we have

$$\tilde{p}\,(\mathbf{Y}|\gamma,c) = (1+c)^{-\frac{q_\gamma+1}{2}} \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\} \exp\left\{-\frac{1}{2(c+1)\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\} \tag{2.9}$$

It is not hard to show this approximation to equation (2.4) has an error of order $O(n^{-1})$ provided the log-likelihood function satisfies certain regularity conditions (See Kass et. al. 1990 for details). That is,

$$p\,(\mathbf{Y}|\gamma,c) = \tilde{p}\,(\mathbf{Y}|\gamma,c)(1 + O(n^{-1}))$$

The proof is given below.

Proof:

Consider Laplace approximation for

$$\begin{aligned} p\,(\mathbf{Y}|\gamma, c) &= \int_{\mathbf{R}^{q_\gamma}+1} p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) p\,(\boldsymbol{\beta}_\gamma|\gamma, c)\, d\boldsymbol{\beta}_\gamma \\ \text{and } \tilde{p}\,(\mathbf{Y}|\gamma, c) &= \int_{\mathbf{R}^{q_\gamma}+1} \tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) p\,(\boldsymbol{\beta}_\gamma|\gamma, c)\, d\boldsymbol{\beta}_\gamma \end{aligned}$$

where $\log \tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$ is the second-order approximation to $\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$ by expanding the later around $\hat{\boldsymbol{\beta}}_\gamma$. Note that $\hat{\boldsymbol{\beta}}_\gamma$ maximize both $\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$ and $\log \tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$, and that they are equal at $\boldsymbol{\beta}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$; also $\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$ and $\log \tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)$ have the same Hessian matrix at $\hat{\boldsymbol{\beta}}_\gamma$. Hence $p\,(\mathbf{Y}|\gamma, c)$ and $\tilde{p}\,(\mathbf{Y}|\gamma, c)$ have the same Laplace approximation denoted $p_L\,(\mathbf{Y}|\gamma, c)$. Therefore,

$$p_L\,(\mathbf{Y}|\gamma, c) = p\,(\mathbf{Y}|\gamma, c)(1 + O(n^{-1}))$$

and

$$p_L\,(\mathbf{Y}|\gamma, c) = \tilde{p}\,(\mathbf{Y}|\gamma, c)(1 + O(n^{-1}))$$

Thus,

$$p\,(\mathbf{Y}|\gamma, c) = \tilde{p}\,(\mathbf{Y}|\gamma, c)(1 + O(n^{-1}))$$

2.2.3 Discussion

1. If $\mathbf{Y}$ is normally distributed, the GLM with the canonical link becomes the familiar normal linear model. As we can easily see, the second-order approximation to the log-likelihood is exactly itself. Hence, in this case, $\tilde{p}\,(\mathbf{Y}|\gamma, c) = p\,(\mathbf{Y}|\gamma, c)$

2. Compare the Integrated Laplace approximation $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$ (equation (2.9)) with the Laplace approximation $p_L\left(\mathbf{Y}|\gamma,c\right)$ (equation (2.6)), we find if we substitute c with $(c+1)$ in $p_L\left(\mathbf{Y}|\gamma,c\right)$, we will have $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$. Both $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$ and $p_L\left(\mathbf{Y}|\gamma,c\right)$ have an error of order $O(n^{-1})$. Raftery (1996) demonstrates empirically that the Laplace approximation is accurate in generalized linear models. So we should expect that the Integrated one also works accurately in practice. However, $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$ is the exact solution under the normal case whereas $p_L\left(\mathbf{Y}|\gamma,c\right)$ is not. Intuitively, $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$ is better than $p_L\left(\mathbf{Y}|\gamma,c\right)$ in the sense that for $p_L\left(\mathbf{Y}|\gamma,c\right)$ we simply replace the prior density function of $\boldsymbol{\beta}_\gamma$ with its evaluation at $\hat{\boldsymbol{\beta}}_\gamma$ then factor it out of integration while for $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$ we directly integrate it out.

3. For large c, $\tilde{p}\left(\mathbf{Y}|\gamma,c\right) \approx p_L\left(\mathbf{Y}|\gamma,c\right)$ for any sample size. For small c, we do not have this at all since:

$$\lim_{c\to 0}\tilde{p}\left(\mathbf{Y}|\gamma,c\right) = \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}$$
$$\cdot\exp\left\{-\frac{1}{2\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\}$$

and

$$\lim_{c\to 0}p_L\left(\mathbf{Y}|\gamma,c\right) \overset{\mathrm{d}=\frac{1}{\mathrm{c}}}{=} \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}$$
$$\cdot\lim_{d\to+\infty}\frac{d^{\frac{q_\gamma+1}{2}}}{\exp\left\{\frac{d}{2\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\}}$$

In the case when $m_\gamma \neq \hat{\boldsymbol{\beta}}_\gamma$, if we apply the L' Hospital's Rule to $\lim_{e\to+\infty}$ by taking the derivative of both the numerator and denominator for finite number of times, we finally arrive at:

$$\lim_{c\to 0} p_L\left(\mathbf{Y}|\gamma, c\right) = 0$$

In the case when $m_\gamma = \hat{\boldsymbol{\beta}}_\gamma$, easily we have

$$\lim_{c\to 0} p_L\left(\mathbf{Y}|\gamma, c\right) = +\infty$$

Now let's look at $p\left(\mathbf{Y}|\gamma, c\right)$ when c equals 0. In this case, $\boldsymbol{\beta}_\gamma$ is fixed at $\mathbf{m}_\gamma$, hence,

$$p\left(\mathbf{Y}|\gamma, c=0\right) = \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma \boldsymbol{m}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma \boldsymbol{m}_\gamma)\cdot \mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi)\cdot \mathbf{1}\right\}$$

Also,we know as $n \to +\infty$, $\hat{\boldsymbol{\beta}}_\gamma \to \mathbf{m}_\gamma$ with probability 1 if the GLM defined in section 2.1 satisfies some regularity assumptions. Therefore, it follows that

$$p\left(\mathbf{Y}|\gamma, c=0\right) = \lim_{n\to+\infty}\lim_{c\to 0}\tilde{p}\left(\mathbf{Y}|\gamma, c\right)$$

In conclusion, when c is small, $\tilde{p}\left(\mathbf{Y}|\gamma, c\right)$ is still a good approximation to $p\left(\mathbf{Y}|\gamma, c\right)$, but $p_L\left(\mathbf{Y}|\gamma, c\right)$ does not.

Based on the above discussions, we have sufficient reason to think $\tilde{p}\left(\mathbf{Y}|\gamma, c\right)$ is better than $p_L\left(\mathbf{Y}|\gamma, c\right)$, and $\tilde{p}\left(\mathbf{Y}|\gamma, c\right)$ is applied as an approximation to $p\left(\mathbf{Y}|\gamma, c\right)$ in later sections.

2.3 The Asymptotic Posterior Mean of $\boldsymbol{\beta}_\gamma$

The posterior distribution of $\boldsymbol{\beta}_\gamma$ is given by

$$p\,(\boldsymbol{\beta}_\gamma|\mathbf{Y},\gamma,c) = \frac{p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma)p\,(\boldsymbol{\beta}_\gamma|\gamma,c)}{p\,(\mathbf{Y}|\gamma,c)} \tag{2.10}$$

Applying the second-order approximation to $\log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma)$ by expanding it around $\hat{\boldsymbol{\beta}}_\gamma$, the posterior distribution of $\boldsymbol{\beta}_\gamma$ can be approximated by a multivariate normal distribution:

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\mathbf{Y},\gamma,c \sim \mathbf{N}_{q_\gamma+1}(\frac{c\,\hat{\boldsymbol{\beta}}_\gamma}{c+1}+\frac{\mathbf{m}_\gamma}{c+1},\frac{c}{c+1}\,\phi\,(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)^{-1}) \tag{2.11}$$

The posterior mean of $\boldsymbol{\beta}_\gamma$ is easily found to be:

$$E(\boldsymbol{\beta}_\gamma|\mathbf{Y},\gamma,c) \simeq \frac{c\,\hat{\boldsymbol{\beta}}_\gamma}{c+1}+\frac{\mathbf{m}_\gamma}{c+1} \tag{2.12}$$

The error order is $O(n^{-1})$. This formula is meaningful in that it tells us that the asymptotic posterior mean of $\boldsymbol{\beta}_\gamma$ is the weighted average of the MLE and the prior mean of $\boldsymbol{\beta}_\gamma$. Note that, $\frac{1}{c+1}$ is a number between 0 and 1; it reflects directly how important the prior mean is, which provides a good reason in later chapters why I re-parameterize c to $k=\frac{1}{c+1}$.

2.4 Calibration

With the priors discussed in section 2.1, I now derive the posterior distribution of γ given hyperparameters c, ω. From equation (1.3),

$$\begin{aligned}
&\pi(\gamma|\mathbf{Y}, c, \omega) \propto \pi(\gamma|\omega) p(\mathbf{Y}|\gamma, c) \\
&= \pi(\gamma|\omega)\tilde{p}(\mathbf{Y}|\gamma, c)(1 + O(n^{-1})) \\
&= \omega^{q_\gamma}(1-\omega)^{p-q_\gamma}(1+c)^{-\frac{q_\gamma+1}{2}} \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi)\cdot\mathbf{1}\right\} \\
&\quad \exp\left\{-\frac{1}{2(c+1)\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\}(1 + O(n^{-1})) \qquad (2.13)
\end{aligned}$$

George and Foster (2000) has shown that for the normal linear model, if we choose proper values of hyperparameters (c, ω), selection criteria such as AIC, C_p, BIC and RIC are equivalent to selection of models with maximum posterior. For GLM, after developing a Bayesian formulation, it is now easy to show by choosing $\mathbf{m}_\gamma, c$ and ω, that the maximum posterior for γ in (2.13) asymptotically corresponds to the model selection criteria AIC and BIC.

To reveal the connection between AIC, BIC and the posterior of γ given c and ω, I re-express $\pi(\gamma|\mathbf{Y}, c, \omega)$ as

$$\begin{aligned}
&\pi(\gamma|\mathbf{Y}, c, \omega) \propto \\
&\quad \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi)\cdot\mathbf{1} - \frac{q_\gamma}{2}\left[2\log\frac{1-\omega}{\omega} + \log(1+c)\right]\right\} \\
&\quad \exp\left\{-\frac{1}{2(c+1)\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\}(1 + O(n^{-1})) \qquad (2.14)
\end{aligned}$$

If we choose $\mathbf{m}_\gamma$ equal to $\hat{\boldsymbol{\beta}}_\gamma$, (2.14) can be simplified to

$$\begin{aligned}\pi\left(\gamma|\mathbf{Y},c,\omega\right) \quad &\propto \quad \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right. \\ &\left. -\frac{q_\gamma}{2}\left[2\log\frac{1-\omega}{\omega} + \log(1+c)\right]\right\}(1+O(n^{-1}))\end{aligned}$$

So when $n\to\infty$, to maximize $\pi\left(\gamma|\mathbf{Y},c,\omega\right)$, we only need to minimize

$$\hat{D}_\gamma + q_\gamma\left(2\log\frac{1-w}{w} + \log(1+c)\right) \tag{2.15}$$

$\hat{D}_\gamma$ is the deviance of model γ,

$$\hat{D}_\gamma \equiv 2\log\hat{L}_s - 2\log\hat{L}_\gamma$$

where $\hat{L}_s$ is the estimated likelihood of the saturated model with n parameters that fits the n observations perfectly. It is easy to see $\hat{L}_s$ is fixed once the data is given. $\hat{L}_\gamma$ is the estimated likelihood under model γ:

$$\hat{L}_\gamma = \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}$$

There are many choices of (c,ω) by which we can calibrate the form (2.15) to selection criteria AIC and BIC. For example,

1. Take $\left\{\begin{array}{lcl} c & = & e^g-1, g \text{ is a constant} \\ w & = & \frac{1}{2}\end{array}\right.$, (2.15) becomes $\hat{D}_\gamma + g\,q_\gamma$. When $g=2$, it is the usual formulation for AIC .
2. Take $\left\{\begin{array}{lcl} c & = & n-1 \\ w & = & \frac{1}{2}\end{array}\right.$, (2.15) becomes $\hat{D}_\gamma + (\log n)\,q_\gamma$ that is BIC .

Hence, with the above (c,ω) combinations, the highest posterior model will correspond exactly to AIC and BIC as $n\to\infty$.

2.5 Specifying the Prior Mean Vector $\mathbf{m}_\gamma$ of $\boldsymbol{\beta}_\gamma$ and the Dispersion Parameter ϕ

In this dissertation, I treat the prior mean vector $\mathbf{m}_\gamma$ of $\boldsymbol{\beta}_\gamma$ and the dispersion parameter ϕ as known constants. In practice, they are unknown for most cases and it is necessary for us to specify reasonable values, no matter whether we use an EB or FB approach in variable selection.

If we choose $\hat{\boldsymbol{\beta}}_\gamma$ as the value of $\mathbf{m}_\gamma$ and the proper values for c and ω, we can calibrate AIC and BIC to the posterior $\pi\left(\gamma|\mathbf{Y}, c, \omega\right)$ asymptotically. This naturally makes us think of choosing $\hat{\boldsymbol{\beta}}_\gamma$ as the value of $\mathbf{m}_\gamma$. However, this is not a wise choice. First of all, it makes our prior on $\boldsymbol{\beta}_\gamma$ heavily dependent on the data (data snooping). Second, it yields a penalty of adding a variable in $\pi(\gamma|\mathbf{Y})$ or $\pi\left(\gamma|\mathbf{Y}, c, \omega\right)$ that is independent of the data, hence can not be adaptively adjusted. Third, as will be shown in Chapter 4, by choosing $\hat{\boldsymbol{\beta}}_\gamma$ as $\mathbf{m}_\gamma$, the penalty in the posterior of models based on uniform hyper-priors is actually a reward of adding a variable if not putting a restriction on c and ω.

Under the situation where strong prior information does not exist, the common choice for $\mathbf{m}_\gamma$ in previous research is $(\bar{\beta}_0, 0, \ldots, 0)$ so that all terms other than the intercept have zero prior means. Reasonably, we can use the MLE of β_0 under the null model as the value of $\bar{\beta}_0$, which is $g(\bar{Y})$ for any link function g or specifically $b'^{-1}(\bar{Y})$ for a canonical link. I follow this in the simulation of this study.

Now let's focus on specifying the value of ϕ. In the exponential family, for some distributions, ϕ is constant, for example, Poisson, Exponential, Bernoulli, Binomial, and Negative Binomial distributions; for other distributions, like Poisson and binomial with overdispersion, or Normal, Gamma, Inverse Gaussian, ϕ is unknown and one may proceed as before with ϕ replaced by an estimate $\hat{\phi}$, as in McCullagh and Nelder (1989). There are three estimates of ϕ summarized by Jorgensen (1987) under model γ:

1. $\hat{\phi}_1 = \frac{D(\mathbf{Y}, \hat{\boldsymbol{\mu}}_\gamma)}{n-q_\gamma-1}$ which is an asymptotic unbiased estimator of ϕ. $D(\mathbf{Y}, \hat{\boldsymbol{\mu}}_\gamma)$ is the deviance for model γ.

2. $\hat{\phi}_2 = \frac{P^2}{n-q_\gamma-1}$, where $P = (\mathbf{Y} - \hat{\boldsymbol{\mu}}_\gamma)V_\gamma^{-1}(\mathbf{Y} - \hat{\boldsymbol{\mu}}_\gamma)$ is the generalized Pearson Statistic. This is actually a moment estimator.

3. $\hat{\phi}_3$ maximizes the following modified profile likelihood for parameter ϕ (Barndorff-Nielsen, 1983): $L^0(\phi) = \phi^{\frac{q_\gamma+1}{2}} p(\mathbf{Y}|\hat{\boldsymbol{\theta}}_\gamma, \phi)$, where $p(\mathbf{Y}|\boldsymbol{\theta}, \phi)$ is the density function of $\mathbf{Y}$.

A reasonable estimate would be any of the above $\hat{\phi}$ under the full model. In McCullagh and Nelder (1989), they used $\hat{\phi}_2$ under the full model as an estimate. For simplicity, I always use the true value of ϕ in the simulation of this study since all the true values of parameters are known from generating the data.

Chapter 3

Empirical Bayes Selection Criteria

There are several possible sources of information about hyperparameters. The most obvious is subjective knowledge from experience or previous research. Often it might be the case that this information is not available as hyperparameters rarely have clear intuitive interpretation in practice. One exception is Random Effects Models (or Mixed Models) where the hyperparameter, the mean of the prior distribution, is meaningful. A second source of possible information is the data itself. One could use the data to estimate hyperparameters of the prior distributions. This type of situation is typically called an Empirical Bayes problem. Extensive development of EB methodology began with Robbins (1955,1964). Many actual applications and references can be found in Berger (1985)[p.169].

In this chapter I describe an EB approach to deal with hyperparameters and generate selection criteria for GLM under the framework I developed in last chapter. Three different EB methods are dicussed here. One is a direct generalization of CML to GLM that was first proposed by George and Foster (2000) for the linear normal regression. The other two, $MCML$ and $HCML$ are based on the CML method but are aimed to improve performance of variable selection.

3.1 Generalization of Criterion CML to GLM

Since c and ω control the expected size and proportion of the nonzero components of $\boldsymbol{\beta}$, it is very important to choose reasonable values for them. As pointed out in George and Foster (2000), the arbitrary selection of c and ω may tend to concentrate the prior away from the true underlying model, especially when p is large. To avoid this, they proposed two empirical Bayes methods that use the data to estimate c and ω under the linear normal setting:

1. Estimators of c and ω are obtained by maximizing the marginal likelihood of c and ω, $L(c, \omega|\mathbf{Y})$, (referred as MML):

   $$L(c, \omega|\mathbf{Y}) \propto \sum_{\gamma} \pi\,(\gamma|\omega) p\,(\mathbf{Y}|\gamma, c)$$

2. Estimator of c and ω are obtained by maximizing the conditional 'likelihood' of c and ω, $L^*(c, \omega, \gamma|\mathbf{Y})$, (referred as CML):

   $$L^*(c, \omega, \gamma|\mathbf{Y}) \propto \pi\,(\gamma|\omega) p\,(\mathbf{Y}|\gamma, c)$$

   This is equivalent to maximizing the largest component of $L(c, \omega|\mathbf{Y})$. Instead of marginalisation over γ that is computationally overwhelming, CML lets the estimators depend on γ.

For GLM, CML can be directly generalized while MML is not feasible since GLM has more computational complexity than linear regression. Conditionally on γ, the estimator of c and ω that maximize L^* when $n \to \infty$ are easily seen to be:

$$\hat{c}_\gamma = \left[\frac{(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T (\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)}{\phi(q_\gamma + 1)} - 1 \right]_+ \tag{3.1}$$

where $(.)_+$ is the positive-part function and

$$\hat{\omega}_\gamma = \frac{q_\gamma}{p} \tag{3.2}$$

Inserting these into the posterior (2.13) and taking the logarithm shows that when $n \to \infty$, the posterior $\pi\left(\gamma | \mathbf{Y}, \hat{c}_\gamma, \hat{\omega}_\gamma\right)$ is maximized by the γ that minimizes

$$C_{CML} = \begin{cases} D_\gamma + (q_\gamma + 1)(\log \frac{T_\gamma}{q_\gamma + 1} + 1) - 2\left\{q_\gamma \log q_\gamma + (p - q_\gamma)\log(p - q_\gamma)\right\} \\ \qquad \text{if } \frac{T_\gamma}{q_\gamma + 1} > 1 \\ D_\gamma + T_\gamma - 2\left\{q_\gamma \log q_\gamma + (p - q_\gamma)\log(p - q_\gamma)\right\} \\ \qquad \text{if } \frac{T_\gamma}{q_\gamma + 1} \le 1 \end{cases} \tag{3.3}$$

where D_γ is the deviance of model γ and $T_\gamma \equiv (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T (\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)/\phi$.

3.2 Two Modified Versions of CML

Notice in (3.2) $\hat{\omega}_\gamma$ does not depend on the data at all. It is actually like deciding the prior probability of including a variable for model γ after looking at how many independent variables model γ has. It may make more sense to choose ω as $\frac{1}{2}$ instead of doing this. So a modified version of CML (referred as $MCML$) is proposed here: only obtain the estimator of c by maximizing the conditional likelihood $L^*(c, \omega, \gamma | \mathbf{Y})$ while fixing ω at a non-informative value

$\frac{1}{2}$. In this case, the estimator of c keeps the exact same form as in (3.1) and the criterion in (3.3) changes to:

$$C_{MCML} = \begin{cases} D_\gamma + (q_\gamma + 1)(\log \frac{T_\gamma}{q_\gamma+1} + 1) & \text{if } \frac{T_\gamma}{q_\gamma+1} > 1 \\ D_\gamma + T_\gamma & \text{if } \frac{T_\gamma}{q_\gamma+1} \leq 1 \end{cases} \tag{3.4}$$

The popular 'non-informative' choice $\omega = \frac{1}{2}$ in $MCML$ yields $\pi(\gamma) \equiv 2^{-p}$ that will concentrate on models with close to $\frac{p}{2}$ nonzero coefficients, which could bring unsatisfactory results when the true model was parsimonious or saturated. To avoid this, we might consider putting a uniform prior on ω directly, that is, $\pi(\omega) = 1$, and integrate it out from $\pi(\gamma|\mathbf{Y}, c, \omega)$, then obtain the estimator of c by maximizing the 'conditional' likelihood of c, $L^*(c, \gamma|\mathbf{Y})$, (referred as $HCML$, i.e., Half-CML):

$$\begin{aligned} L^*(c, \gamma|\mathbf{Y}) &\propto \int_0^1 \pi(\gamma|\omega) p(\mathbf{Y}|\gamma, c) \pi(\omega) d\omega \\ &= \int_0^1 \omega^{q_\gamma} (1-\omega)^{p-q_\gamma} d\omega \, p(\mathbf{Y}|\gamma, c) \\ &= \frac{\Gamma(q_\gamma + 1)\Gamma(p - q_\gamma + 1)}{\Gamma(p+2)} p(\mathbf{Y}|\gamma, c) \end{aligned} \tag{3.5}$$

From (3.5), we can easily see the estimator of c keeps the same form as before and the criterion changes to:

$$C_{HCML} = \begin{cases} D_\gamma + (q_\gamma + 1)(\log \frac{T_\gamma}{q_\gamma+1} + 1) - 2\left\{\log\Gamma(q_\gamma+1) + \log\Gamma(p - q_\gamma + 1)\right\} \\ \qquad \text{if } \frac{T_\gamma}{q_\gamma+1} > 1 \\ D_\gamma + T_\gamma - 2\left\{\log\Gamma(q_\gamma+1) + \log\Gamma(p - q_\gamma + 1)\right\} \\ \qquad \text{if } \frac{T_\gamma}{q_\gamma+1} \leq 1 \end{cases} \tag{3.6}$$

3.3 Discussion

The EB estimates of hyperparameters may be criticized for its dependence on the data. One may feel more comfortable with a model proposed on intuitive or theoretical grounds as opposed to one dependent on "data snooping". However, there are several reasons that the EB criteria are considered in this study:

1. The EB methods avoid the difficulties of choosing priors on hyperparameters. The EB criteria, as shown above, are easily computed and all have asymptotically closed forms that can be nicely interpreted. When the sample size n is reasonably large, they can be applied conveniently.

2. The EB methods allow for a type of frequentist justification (see Morris, 1983a,b), making them more attractive to non-Bayesians.

3. Some theoretical literature (see Berger, 1985, p.169) have shown that the EB mothods have the property of asymptotic optimality, which means an EB procedure is asymptotically optimal if, as the dimension $p \to \infty$, the procedure is as good (in some sense) as the Bayes procedure were the prior actually known.

4. Although some FB procedure should dominate an EB procedure, how to do this has not been shown in variable selection problems.

The performance of CML, $MCML$ and $HCML$ is evaluated, and compared by simulation with AIC, BIC and FB criteria in Chapter 7.

Chapter 4

Fully Bayes Selection Criteria

A Fully Bayes approach entails putting hyperpriors on c and ω and then integrating them out to compute the posterior $\pi(\gamma|\mathbf{Y})$, instead of choosing c and ω based on the data; that is:

$$\pi(\gamma|\mathbf{Y}) \propto \iint_{c,\omega} p(\mathbf{Y}|\gamma, c)\, \pi(\gamma|\omega)\, \pi(c)\, \pi(\omega)\, dc\, d\omega \tag{4.1}$$

where $\pi(\gamma|\omega)$ is specified by (2.1), $\pi(\omega)$ and $\pi(c)$ are hyperpriors on ω and c respectively. There are several advantages of the FB approach:

1. It incorporates the hyperparameter estimation error in the analysis while the EB approach does not.

2. One can incorporate actual subjective prior information at the second stage with only slight additional difficulty.

3. One can easily incorporate further unknown parameters like ϕ into the analysis.

4. There exists an admissible FB solution with respect to some priors while the EB criteria are not admissible.

4.1 Re-parameterization of Hyperparamter c

The Integrated Laplace approximation $\tilde{p}\left(\mathbf{Y}|\gamma, c\right)$ to $p\left(\mathbf{Y}|\gamma, c\right)$ given in formula (2.9) is actually a function of $\frac{1}{c+1}$. Noticing that c belongs to $[0, +\infty)$ will lead to an improper uniform prior on c, we would like to re-parameterize c to k by defining $k \equiv \frac{1}{c+1}$ that makes k belong to $(0, 1]$, which gives us a proper uniform prior on k. As discussed in section 2.3, k is actually the prior probability we assign to the prior mean against the MLE of $\boldsymbol{\beta}_\gamma$ when calculating the Bayesian updated posterior mean of $\boldsymbol{\beta}_\gamma$. So it reflects directly how important we think the prior mean of $\boldsymbol{\beta}_\gamma$ is. Also, this re-parameterization enables us to determine more easily what the conjugate hyperpriors are.

If we take a uniform prior in (0,1] on k, we know that

$$f_k(k) = 1$$

$$dk = -\frac{1}{(c+1)^2}\,dc$$

Based on the one-to-one transformation theory of a random variable, the density function of c is

$$f_c(c) = \frac{1}{(c+1)^2}$$

Generally for any density function $f_k(k)$ of k, we have the relationship:

$$f_c(c) = \frac{f_k(k(c))}{(c+1)^2} \tag{4.2}$$

where $k(c)$ is $\frac{1}{c+1}$.

Under this transformation and from equation (4.1), we have

$$\pi(\gamma|\mathbf{Y}) \propto \iint_{k,\omega} p(\mathbf{Y}|\gamma,k)\,\pi(\gamma|\omega)\,\pi(k)\,\pi(\omega)\,dk\,d\omega \tag{4.3}$$

We can easily go from (4.3) back to (4.1) by applying formula (4.2).

Next, I show by choosing either noninformative or conjugate hyperpriors on k and ω, a closed-form posterior $\pi(\gamma|\mathbf{Y})$ can be obtained when $n \to \infty$. Also, to improve the performance of variable selection, I propose a method that uses a restricted integration region on (k,ω) to obtain $\pi(\gamma|\mathbf{Y})$ based on a practical view of getting rid of the disturbance of the full model.

4.2 Model Posterior Based on "Noninformative" Hyperpriors

If we do not have any meaningful prior knowledge about k and ω, it is natural to choose the uniform distribution on [0,1]. From (4.3), we have the posterior distribution of γ if $\mathbf{m}_\gamma \neq \hat{\boldsymbol{\beta}}_\gamma$.

$$\begin{aligned}
\pi\left(\gamma|\mathbf{Y}\right) \propto & \int_0^1\!\!\int_0^1 \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} k^{\frac{q_\gamma+1}{2}} \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right] \\
& \exp\left[-\frac{k}{2\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right] d\omega dk\cdot(1+O(n^{-1})) \\
= \quad & \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}\Gamma(\frac{q_\gamma+3}{2})\left[\frac{(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)}{2\phi}\right]^{-\frac{q_\gamma+3}{2}} \\
& \cdot G_{\alpha=\frac{q_\gamma+3}{2},\beta=1}\left(\frac{1}{2\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right) \\
& \cdot\exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right](1+O(n^{-1}))
\end{aligned} \tag{4.4}$$

where $G_{\alpha=\frac{q_\gamma+3}{2},\beta=1}(.)$ is the CDF of Gamma($\alpha = \frac{q_\gamma+3}{2}, \beta = 1$).

Let's have a close look at this asymptotic posterior of γ. After taking logarithm and throwing away the constant part for all models, we can decompose it into three parts: the first part E_ω is related to the integration over ω, given by

$$E_\omega = \log\Gamma\left(q_\gamma + 1\right) + \log\Gamma(p - q_\gamma + 1)$$

the second part E_k is related to the integration over k or equivalently c, given by

$$E_k = \log\Gamma\left(\frac{q_\gamma + 3}{2}\right) - \frac{q_\gamma + 3}{2}\log\frac{T_\gamma}{2} - \log G_{\alpha=\frac{q_\gamma+3}{2},\beta=1}\left(\frac{T_\gamma}{2}\right)$$

where $T_\gamma \equiv (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)/\phi$; the third part is actually the estimated log-likelihood of model γ, $\log\hat{L}_\gamma$, which is always increasing as q_γ goes up. It is easy to see that E_ω is a convex function of q_γ. As q_γ increases, it is decreasing first until q_γ reaches $[\frac{p-1}{2}]$ and then increasing afterwards. And E_ω for the full model is equal to E_ω for the null model. So if we ignore the effect of the integration over k, E_ω will always make us choose the full model since $\log\hat{L}(Full) > \log\hat{L}(Null)$. Hence, E_k plays a very important role in selecting models. Actually, E_k is the only part in $\pi(\gamma|\mathbf{Y})$ that has connection with the data through T_γ and it should penalize a model with a large q_γ. Let's take $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}$ under the null model, i.e., $(\bar{\beta}_0, 0, \ldots, 0)$ and see how E_k works. Now we could think of $T_\gamma = (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)/\phi$ as a measure of the distance between the null model and model γ. A model with a large number of nonzero components tends to have a large distance from the null model hence a large value of T_γ. We know $\log\frac{T_\gamma}{2}$ and $\log G(\frac{T_\gamma}{2})$ are both increasing functions of T_γ, so E_k penalizes a model with a large q_γ by reversing the sign of both.

4.3 Model Posterior Based on Noninformative Hyperpriors with Restricted Integration Region

In practice, people tend to consider a large set of potential predictors or explanatory variables when they build regression models, with the hope that no important ones are left out, and resort to the variable selection algorithms to choose the "best" model for them. This leads to the consideration of many redundant variables where the full model is unlikely to be the "best" one, especially when p is large. However, even with the no-preference priors, there are two parts E_ω and $\log \hat{L}_\gamma$, and the first term $\log \Gamma(\frac{q_\gamma+3}{2})$ in E_k that always have largest values under the full model in the posterior of γ given in (4.4). Hence, we have found that the FB selection criterion often wrongly chooses the full model as the "best" one, as confirmed by the simulation results. This will make the FB criterion work poorly sometimes, especially when the null model is the true model or the model is close to the full model .

To overcome this difficulty, I go back to $\pi(\gamma|\mathbf{Y}, c, \omega)$ given in (2.14) and look for hints as to why $\pi(\gamma|\mathbf{Y})$ tends to put high probability on the full model. We can easily find under the case of taking $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$ in (2.14), as AIC and BIC do, that if $2\log\frac{1-\omega}{\omega} + \log(c+1) < 0$, adding variables to a model will increase $\pi(\gamma|\mathbf{Y}, c, \omega)$ anyway; as a result, for any c and ω that satisfies $2\log\frac{1-\omega}{\omega} + \log(c+1) < 0$, the highest posterior model will always be the full model. For other choices of m_γ, this will make the term $\exp\left\{-\frac{q_\gamma}{2}[2\log\frac{1-\omega}{\omega} + \log(c+1) < 0]\right\}$ in (2.14) in effect reward adding a variable. To avoid this, we may restrict $2\log\frac{1-\omega}{\omega} + \log(c+1) \geq 0$. By doing so, under the uniform priors on k and ω and $T_\gamma \neq 0$ (that is, $\mathbf{m}_\gamma \neq \hat{\boldsymbol{\beta}}_\gamma$), we have (calculation details in appendix B)

$$\pi(\gamma|\mathbf{Y}) \propto \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right]\cdot\Gamma\left(\frac{q_\gamma+3}{2}\right)$$
$$\cdot\left(\frac{T_\gamma}{2}\right)^{-\frac{q_\gamma+3}{2}}\cdot\left\{\frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}B_{(q_\gamma+1,p-q_\gamma+1)}\left(\frac{1}{2}\right)\cdot G_{(\frac{q_\gamma+3}{2},1)}\left(\frac{T_\gamma}{2}\right)\right.$$
$$\left.+\int_{0.5}^{1}\omega^{q_\gamma}(1-\omega)^{p-q_\gamma}\cdot G_{(\frac{q_\gamma+3}{2},1)}\left(\left(\frac{T_\gamma}{2}\right)\left(\frac{1}{\omega}-1\right)^2\right)d\omega\right\}(1+O(n^{-1}))\quad(4.5)$$

where $B_{q_\gamma+1,p-q_\gamma+1}(.)$ is the CDF of $Beta(\alpha = q_\gamma+1, \beta = p-q_\gamma+1)$. There is still an integration left in equation (4.5). It is not hard to evaluate it through simple numerical methods. However, the complicated form of (4.5) makes it hard to see how the restriction on k and ω penalizes the full model or a model with a large q_γ. For illustration, let's take $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$ that makes the penalty under the uniform priors on k and ω have a simple form and have no connection with the data, so that the effect of the restriction can be analyzed. Under the uniform priors with $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$, without restricting k and ω, from (4.4) we have

$$\pi(\gamma|\mathbf{Y}) \propto \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}\cdot\frac{2}{q_\gamma+3}$$
$$\cdot\exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right](1+O(n^{-1}))\quad(4.6)$$

With the restriction $2\log\frac{1-\omega}{\omega} - \log k \geq 0$, we have (calculation details in appendix B)

$$\pi\left(\gamma|\mathbf{Y}\right) \propto$$
$$\frac{2}{q_\gamma+3}\left[\frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}B_{q_\gamma+1,p-q_\gamma+1}\left(\frac{1}{2}\right)+\int_{\frac{1}{2}}^{1}\omega^{-3}(1-\omega)^{p+3}d\omega\right]$$
$$\cdot\exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-\mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi}+\mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right](1+O(n^{-1})) \tag{4.7}$$

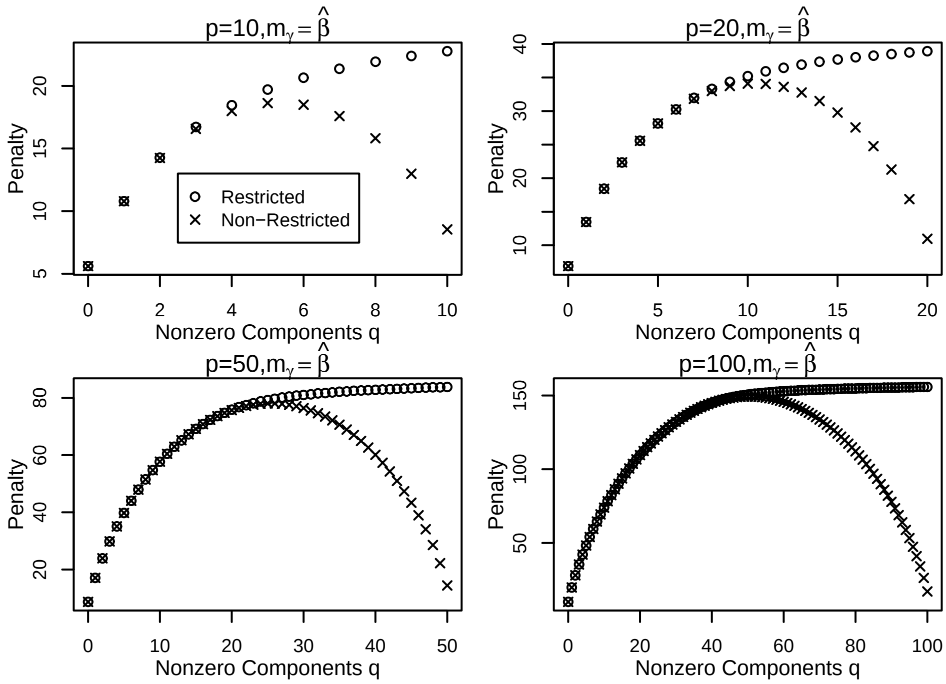

Figure 4.1: The Effect of the Restriction $2\log\frac{1-\omega}{\omega}-\log k\geq 0$

After taking -2 times of its logarithm, we can decompose (4.6) and (4.7) into two parts: the first part is the penalty and the second part $-2\log\hat{L}_\gamma$ is equivalent to the deviance of model γ. To see how the restriction works, I plot the penalty parts for $\pi\left(\gamma|\mathbf{Y}\right)$ both with and without the restriction in

Figure 4.1. Clearly, the penalty without restriction is a concave function and it puts large values around $\frac{p}{2}$ and small values around 0 or p; the penalty with restriction is an increasing function of q_γ and the maximum value is put at $q_\gamma = p$. Hence the effect of the restriction is to increase the penalty on models with large q_γ.

4.4 Model Posterior Based on Conjugate Hyperpriors

It is obvious that the conjugate prior distribution for k is the truncated Gamma distribution since k can not be greater than 1; the conjugate prior distribution for ω is the Beta distribution. Let's assume

$$\begin{aligned} \omega &\sim Beta(\alpha, \beta) \\ k &\sim Truncated\ Gamma(a, b), k \in (0,1) \end{aligned}$$

Therefore, without restricting k and ω, from (26) if $\mathbf{m}_\gamma \neq \hat{\boldsymbol{\beta}}_\gamma$ we have,

$$\begin{aligned} \pi\left(\gamma|\mathbf{Y}\right) \propto & \frac{\Gamma(\alpha+q_\gamma)\Gamma(\beta+p-q_\gamma)}{\Gamma(\alpha+\beta+p)}\Gamma\left(\frac{q_\gamma+1}{2}+a\right) \\ & \cdot \left(\frac{T_\gamma}{2}+\frac{1}{b}\right)^{-\frac{q_\gamma+1}{2}-a} G_{\left(\frac{q_\gamma+1}{2}+a,1\right)}\left(\frac{T_\gamma}{2}+\frac{1}{b}\right) \\ & \cdot \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right](1+O(n^{-1})) \end{aligned} \tag{4.8}$$

With restricting k and ω by $2\log\frac{1-\omega}{\omega} - \log k \geq 0$ and assuming $T_\gamma + \frac{1}{b} \neq 0$, we have (calculation details in appendix B)

$$\begin{aligned}
&\pi\left(\gamma|\mathbf{Y}\right) \propto \\
&\quad \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right] \\
&\quad \cdot\Gamma\left(\frac{q_\gamma+1}{2}+a\right)\cdot\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)^{-\frac{q_\gamma+1}{2}-a}\cdot\left\{\frac{\Gamma(q_\gamma+\alpha)\Gamma(p-q_\gamma+\beta)}{\Gamma(p+\alpha+\beta)}\right. \\
&\quad B_{(q_\gamma+\alpha,p-q_\gamma+\beta)}(0.5)\cdot G_{\left(\frac{q_\gamma+1}{2}+a,1\right)}\left(\frac{T_\gamma}{2}+\frac{1}{b}\right) \\
&\quad \left.+\int_{0.5}^{1}\omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1}\cdot G_{\left(\frac{q_\gamma+1}{2}+a,1\right)}\left(\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)\left(\frac{1}{\omega}-1\right)^2\right)d\omega\right\} \\
&\quad (1+O(n^{-1}))
\end{aligned} \tag{4.9}$$

Through the conjugate priors, theoretically we can incorporate actual subjective prior information at the second stage. For example, for ω's prior, if we choose Beta(0.5, 0.5) as the prior distribution, we actually put more weight on ω values close to 0 and 1 since the PDF is bathtub-shaped; if we choose Beta(1.5, 1.5), we put more weight on ω values close to 0.5 since the PDF is concave; if we choose Beta(2, 1), we put more weight on large ω values since the PDF is a line with a positive slope; if we choose Beta(1, 2), we put more weight on small ω values since the PDF is a line with a negative slope. For the prior on k, it is suggested in the literature (see Zellner 1986, Smith & Kohn 1996) that c be chosen large, which is obtained by small k. So we might consider a special form $f_k(k) = (1-\rho)k^{-\rho}$,$0 < \rho < 1$ that put more weight on small k's. This prior is actually truncated Gamma(1-ρ, ∞).

It is worth mentioning that the noninformative hyperpriors on k and ω discussed in section 4.2 and 4.3 are actually a special case of the conjugate hyperpriors, that is, a=1,b=$+\infty$, $\alpha = 1$ and $\beta = 1$.

Chapter 5

Conditional Fully Bayes Criteria

In this chapter, I describe an alternative Fully Bayes approach based on a different hyper-parameterization. Recall that I have introduced two hyper-parameters c and ω through the priors on the first-stage parameters. Through its influence on the prior covariance matrix of $\boldsymbol{\beta}_\gamma$, the hyperparameter c controls the expected size of the nonzero coefficients of $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$. Under the model prior (2.1), the components of $\boldsymbol{\beta}$ are i.i.d. nonzero with probability ω, so the hyperparameter ω controls the expected proportion of such nonzero components. Now let me make a step further, relaxing c to c_γ, that is, I take the prior distribution on $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ as below.

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\gamma, c_\gamma \sim \mathbf{N}_{q_\gamma+1}(\mathbf{m}_\gamma, c_\gamma\,\phi\,(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)^{-1}) \text{ for } c_\gamma \in (0, +\infty) \tag{5.1}$$

By doing so, we have model specific hyperparameters c_γ instead of an overall c for all 2^p models. This means for each model in the model space, there is a hyperparameter that controls the expected size of the corresponding β coefficients.

The idea of relaxing c to c_γ is motivated by observing CML has good performance in variable selection under the normal linear setting in George and Foster (2000). As we already know, CML belongs to the EB methods and it should estimate c from the data. Due to the difficulty caused by the summation over γ in $L(c, \omega|\mathbf{Y})$, CML gives a heuristic approach by letting the estimators of c depend on γ. This suggests relaxing c to c_γ in the prior of $\boldsymbol{\beta}_\gamma$. The benefit of doing this is to make CML no longer a heuristic EB method: the estimators of c become the true MLE of c_γ based on the data. Since CML has been reported to work well in George and Foster (2000), it is interesting to consider an FB approach for c_γ, in which adjustable proper hyperpriors that depend on γ can be chosen for c_γ. I call this FB approach as a "Conditional Fully Bayes" or in brief, CFB approach.

One may question whether all the formulas given in the Bayesian framework are still valid for CFB as I let the hyperparameter c dependent on γ. It would be better to check this before I proceed.

First, what happens to the marginal distribution of the data $\mathbf{Y}$ and the model posterior given hyperparameters? Instead of writing $p(\mathbf{Y}|\gamma, c)$ and $\pi(\gamma|\mathbf{Y}, c, \omega)$, we should write $p(\mathbf{Y}|\gamma, c_\gamma)$ and $\pi(\gamma|\mathbf{Y}, c_\gamma, \omega)$, respectively. To obtain both, we treat c_γ as a constant and integrate the model specific parameters $\boldsymbol{\beta}_\gamma$ out. It might be clearer to think of vector $\mathbf{c} = (c_1, \ldots, c_\gamma, \ldots, c_{2^p})$ as a constant vector at this stage. No matter which model one looks at, c_γ is a constant. So it is obvious that everything goes exactly the same as before except we replace c with c_γ in any formulas regarding to $p(\mathbf{Y}|\gamma, c)$ and $\pi(\gamma|\mathbf{Y}, c, \omega)$, for example, formula (2.9), (2.13).

Second, what about the model posterior $\pi(\gamma|\mathbf{Y})$? To obtain this, we have to integrate all the hyperparameters out of $\pi(\gamma|\mathbf{Y}, c_\gamma, \omega)$, that is,

$$\pi(\gamma|\mathbf{Y}) \propto \iint_{\mathbf{c},\,\omega} p(\mathbf{Y}|\gamma, c_\gamma)\,\pi(\gamma|\omega)\,\pi(\mathbf{c})\,\pi(\omega)\,d\mathbf{c}\,d\omega \tag{5.2}$$

Note in (5.2) that, $\mathbf{c}$ is a hyperparameter vector $(c_1, \ldots, c_\gamma, \ldots, c_{2^p})$. It is easy to get

$$\pi(\gamma|\mathbf{Y}) \propto \iint_{c_\gamma,\,\omega} p(\mathbf{Y}|\gamma, c_\gamma)\,\pi(\gamma|\omega)\,\pi(c_\gamma)\,\pi(\omega)\,dc_\gamma\,d\omega \tag{5.3}$$

Now it is clear that the Bayesian framework has not been changed and it is reasonable for us to use all the formulas developed in Chapter 2. However, the CFB approach is theoretically more complicated than the FB approach discussed in last chapter in a sense that CFB has $2^p + 1$ hyperparameters while FB only has two hyperparameters. It is impossible for us to address hyperpriors for c_γ individually, so great efforts are required in choosing them systematically. In this chapter, a workable solution is presented.

5.1 Re-parameterization of Hyperparameters c_γ

Recall that I reparameterized hyperparameter c to k in the FB approach. Based on the same reasons discussed in Chapter 4, I now do this to each c_γ, but with a more general transformation form.

Define a new hyperparameter k_γ for each γ as

$$k_\gamma \equiv \frac{1}{(c_\gamma + 1)^\tau}$$

where $\tau(.)$ is a positive function of model γ, that is $\tau(\gamma) > 0$. Obviously, $k_\gamma \in (0, 1]$.

Now it is convenient for us to consider an uniform hyperprior on k_γ for any γ. Based on the the one-to-one transformation theory of a random variable, $f_{c_\gamma}(c_\gamma)$, the density function of c_γ, is as below.

$$f_{c_\gamma}(c_\gamma) = \frac{\tau(\gamma)}{(c_\gamma + 1)^{\tau(\gamma)+1}} \tag{5.4}$$

More generally, the relationship between the two density functions for any hyperprior on k_γ is

$$f_{c_\gamma}(c_\gamma) = \frac{\tau(\gamma)\, f_{k_\gamma}(k_\gamma(c_\gamma))}{(c_\gamma + 1)^{\tau(\gamma)+1}} \tag{5.5}$$

where $k_\gamma(c_\gamma)$ is $\frac{1}{(c_\gamma+1)^\tau}$.

It is easy to see that we could put the same hyperpriors on k_γ for different γ but actually obtain adjustable hyperpriors on c_γ through the function $\tau(\gamma)$. Therefore by selecting a proper function $\tau(.)$, we may overcome the difficulty of choosing hyperpriors for all the 2^p c_γ hyperparameters individually. Here, $\tau(.)$ is called the knob function. By choosing different knob functions, we could obtain different selection criteria that reflect different prior information. I will discuss how I choose the knob function later in this chapter.

5.2 Model Posterior Based on "Noninformative" Hyperpriors

Generally, subjective information about hyperparameters is unlikely available. As in the FB approach, one might feel comfortable to take noninformative hyperpriors on both k_γ and ω. Since after reparameterization k_γ is in $(0,1]$ for any γ, it is natural to put uniform priors on both k_γ and ω, then calculate the model posterior. If $\mathbf{m}_\gamma \neq \hat{\boldsymbol{\beta}}_\gamma$, we have

$$\begin{aligned}
\pi(\gamma|\mathbf{Y}) \propto & \int\int_{k_\gamma\, \omega} p(\mathbf{Y}|\gamma,k_\gamma)\pi(\gamma|\omega)\pi(k_\gamma)\pi(\omega)dk_\gamma d\omega \\
= & \int_0^1\int_0^1 k_\gamma^{\frac{q_\gamma+1}{2\tau}} \exp\left\{-\frac{1}{2}k_\gamma^{\frac{1}{\tau}}\cdot T_\gamma\right\} \\
& \cdot\exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot 1}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot 1\right\} \\
& \cdot\omega^{q_\gamma}(1-\omega)^{p-q_\gamma}dk_\gamma d\omega\,(1+O(n^{-1})) \\
= & \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}\cdot\tau\cdot\Gamma\left(\frac{q_\gamma+1}{2}+\tau\right)\cdot\left(\frac{T_\gamma}{2}\right)^{-\frac{q_\gamma+1}{2}-\tau} \\
& \cdot G_{\alpha=\frac{q_\gamma+1}{2}+\tau,\beta=1}\left(\frac{T_\gamma}{2}\right) \\
& \cdot\exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot 1}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot 1\right]\cdot(1+O(n^{-1})) \quad (5.6)
\end{aligned}$$

where $T_\gamma = (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)/\phi$.

If $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$, $\pi(\gamma|\mathbf{Y})$ is simplified:

$$\begin{aligned}
\pi(\gamma|\mathbf{Y}) \propto & \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)}\cdot\tau\cdot\frac{1}{\frac{q_\gamma+1}{2}+\tau} \\
& \cdot\exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot \mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot \mathbf{1}\right](1+O(n^{-1})) \quad (5.7)
\end{aligned}$$

Unlike the FB approach, $\pi(\gamma|\mathbf{Y})$ has not been ready for use in variable selection as the knob function $\tau(\gamma)$ is necessary. This may be considered both good and bad. Good, because $\pi(\gamma|\mathbf{Y})$ can be much more flexible through adjusting the knob function. For example, if we take $\tau(\gamma)=1$, (5.6) will reduce to (4.4) for the FB approach. We know the FB criteria without restricting the integration area tend to put high probability on the full model hence do not work well sometimes. CFB brings an alternative way to fix this problem by adjusting $\tau(\gamma)$, which will have computational simplicity compared to the FB criteria with restriction. Bad, because there are numerous choices of $\tau(\gamma)$ and it would be better for any decisions to have a theoretical justification. One could develop an intuitive theory to choose the knob function. Certainly, this is not an easy job. To avoid this difficulty, I use some rough guidelines and graphical information to choose it.

5.3 Choosing the Knob Function τ

The knob function is a function of model γ, or more realistically, the number of nonzero components q_γ in model γ. From now on, let's write $\tau(\gamma)$ as $\tau(q_\gamma)$. Even we could incorporate the sample size n and the number of potential covariates p into the knob function. Once data are given, both n and p are treated as constants, so we simply write $\tau(q_\gamma)$ instead of $\tau(q_\gamma, n, p)$.

Though seemingly very different between (5.7) and AIC or BIC, $\pi\left(\gamma|\mathbf{Y}\right)$ in (5.7) can be tuned to mimic the performance of AIC and BIC by selecting appropriate knob functions. In (5.7), if we choose

$$\tau(q_\gamma) = \frac{q_\gamma + 1}{2\,\Gamma(q_\gamma + 1) \cdot \Gamma(p - q_\gamma + 1) \cdot \exp(q_\gamma)} \tag{5.8}$$

$\pi\left(\gamma|\mathbf{Y}\right)$ is equivalent to AIC as $n \to \infty$. Compared to $\frac{q_\gamma+1}{2}$, $\tau(q_\gamma)$ here is a function with very small positive values as $p > 4$. Hence, $\frac{q_\gamma+1}{2} \cdot \frac{1}{\frac{q_\gamma+1}{2}+\tau} \simeq 1$

and (5.7) becomes:

$$\pi(\gamma|\mathbf{Y}) \propto \exp(-q_\gamma) \cdot \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right](1+O(n^{-1})) \quad (5.9)$$

It is easy to see from (5.9), $\pi(\gamma|\mathbf{Y})$ is asymptotically equivalent to AIC.

Similarly in (5.7), if we choose

$$\tau(q_\gamma) = \frac{q_\gamma + 1}{2\,\Gamma(q_\gamma+1)\cdot\,\Gamma(p-q_\gamma+1)\cdot\,n^{\frac{q_\gamma}{2}}} \quad (5.10)$$

$\pi(\gamma|\mathbf{Y})$ is equivalent to BIC as $n \to \infty$. Also, if we choose

$$\tau(q_\gamma) = \frac{q_\gamma + 1}{2\,\Gamma(q_\gamma+1)\cdot\,\Gamma(p-q_\gamma+1)\cdot\,p^{q_\gamma}} \quad (5.11)$$

$\pi(\gamma|\mathbf{Y})$ is equivalent to RIC that is a selection criterion for linear regression as $n \to \infty$.

One could imagine that variable selection criteria with better performance than AIC and BIC be obtained under some proper knob functions. If we take $\tau(q_\gamma) = 1$, $\pi(\gamma|\mathbf{Y})$ becomes the FB criterion without restriction in (4.6), which does not penalize a model with a large q_γ enough (note when I mention the FB criterion, without explicitly stating it, I always refer to the one under "noninformative" hyperpriors). We could avoid this in CFB by doing so: first consider $\tau(q_\gamma)$ be a function with very small values compared to $\frac{q_\gamma+1}{2}$, which makes $\frac{q_\gamma+1}{2}\cdot\frac{1}{\frac{q_\gamma+1}{2}+\tau} \simeq 1$ hold. Then the major difference between $\pi(\gamma|\mathbf{Y})$ for CFB and $\pi(\gamma|\mathbf{Y})$ for FB without restriction is that CFB has a multiplier τ. So we could choose $\tau(q_\gamma)$ as a decreasing function of q_γ, which penalizes a model with a large q_γ. For example, $\tau_1(q_\gamma) = \frac{1}{q_\gamma+1}$ is the simplest function that satisfies the above guidelines. Other choices include $\tau_2(q_\gamma) = p^{-q_\gamma}$, $\tau_3(q_\gamma) = \exp(-q_\gamma)$, $\tau_4(q_\gamma) = \frac{1}{\Gamma(q_\gamma+1)}$ and so on. For illustration, I take $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$ and plot the penalty part in (5.7) for the above four τ functions, which is $-2\log\Gamma(q_\gamma+1) - 2\log\Gamma(p-q_\gamma+1) + 2\log\Gamma(p+2) - 2\log(\tau) + 2\log(\frac{q_\gamma+1}{2}+\tau)$. Along with them, the penalties for AIC which is $2\,q_\gamma$, RIC which is $2\,\log(p)\cdot q_\gamma$ and the data-independent penalty for the FB criterion with restriction in (4.7) are also given in Figure 5.1.

Further in (5.6), we can incorporate data information into the penalty part by taking $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}$ under the null model. By bringing in a data-dependent penalty part through T_γ, the performance of selection criteria would be improved when the true model has a small number of nonzero components but worse when the true model is close to full, especially when we put large data-independent penalty on the full model already.

Through a preliminary simulation, I have observed that AIC tends to pick up large models so works poorly especially when the true model has a small number of nonzero components, which indicates the penalty of AIC does not have a large enough slope when q_γ is small. And $FB.r$, the FB criterion with restriction of integration area, works well compared to other criteria for all q_γ except for $q_\gamma = p$; it indicates that if the data-independent penalty curve of $FB.r$ goes down a little bit when q_γ is close to p, it would achieve better performance when the true model is full. So we might guess that a proper penalty curve should have a larger positive slope at small q than AIC, and go down a bit when q_γ is close to p. From Figure 5.1, we can see the curve of $\tau_1(q_\gamma) = \frac{1}{q_\gamma+1}$ mimics the curve of $FB.r$ closely when q_γ is small but goes down a bit when q_γ is close to p. So we would expect that $CFB.\tau_1$ would work similarly to $FB.r$ but improve at the full model. The curve of $\tau_2(q_\gamma) = p^{-q_\gamma}$ has a large positive slope as RIC at small q_γ, but does not go down when q_γ is close to p, so we would expect it would have very good performance when the true model is small, but not so good near the full model. For the other two functions, the performance is intermediate, not as good as τ_2 at small q_γ but not so bad at the full model. Note here I am talking about the expected performance of CFB criteria with data-dependent penalties in which case $\mathbf{m}_\gamma$ is taken as the MLE of $\boldsymbol{\beta}$ under the null model, though in the graph I can only show the data-independent part in which case $\mathbf{m}_\gamma$ is taken as $\hat{\boldsymbol{\beta}}_\gamma$.

It is also worth mentioning that through a pilot simulation, I have an impression that the performance of CFB selection criteria is pretty robust to the choices of the knob function under the rough guidelines. I tried many functions and most times I got much better performance at small or medium q_γ than AIC and BIC, but comparable or a bit worse performance when q_γ is close to p than BIC.

In the simulation of this study, I evaluate the performance of the CFB criterion with knob function τ_1.

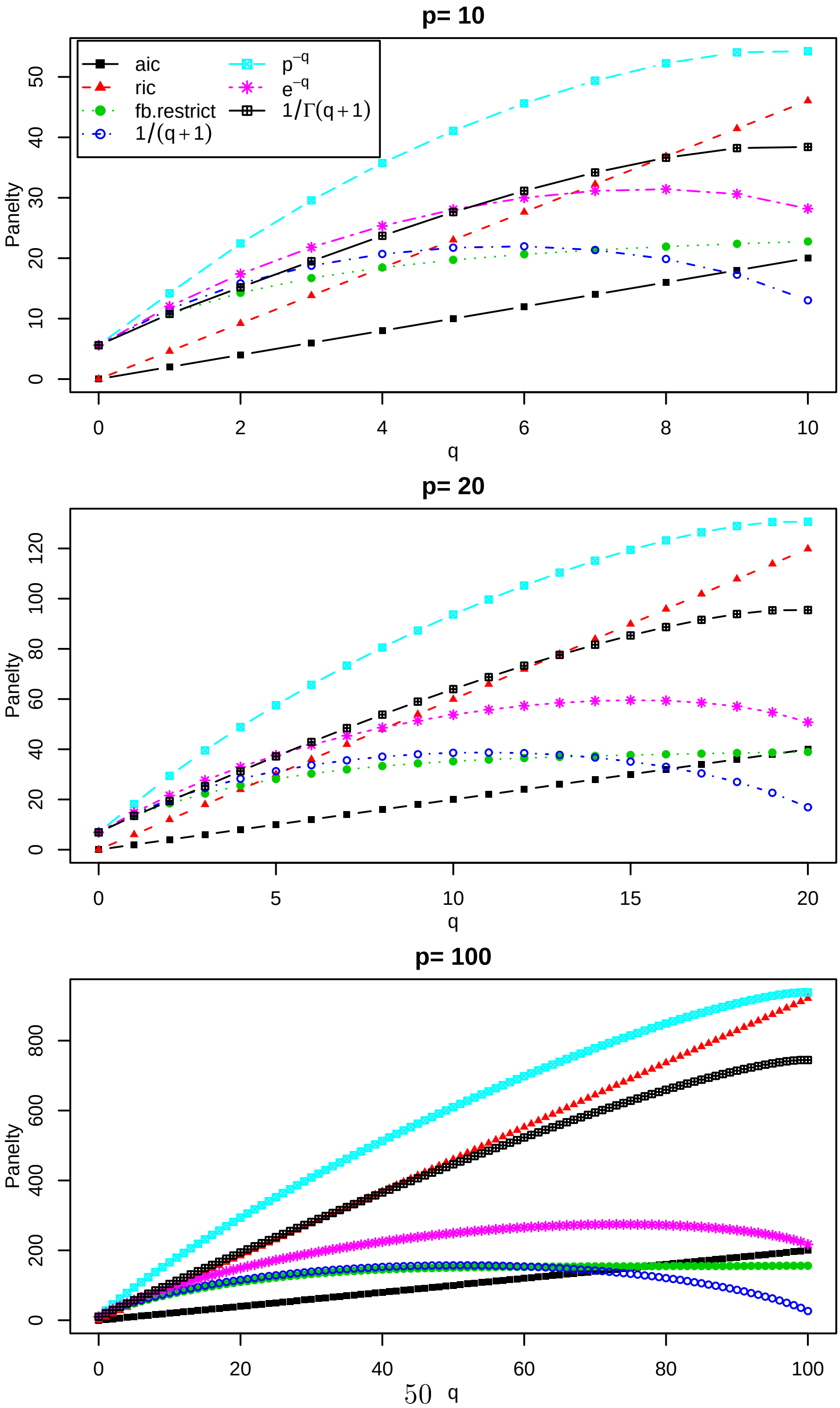

Figure 5.1: The Penalties for Different Knob Functions τ

5.4 Discussion

I like this CFB approach for several reasons.

1. Though CFB has $2^p + 1$ hyperparameters, this does not bring me many difficulties in choosing hyperpriors.

2. The CFB criteria can be very flexible by adjusting the knob function. It can be tuned to classic criteria or the FB criteria.

3. CFB has no numerical integral left in model posterior hence has a simpler form than the FB criterion with restriction. Its model posterior can be better understood.

4. CFB has its corresponding EB approach, CML, while CML is only a heuristic EB method corresponding to the FB approach.

5. My pilot simulation shows that, without much effort in choosing the knob function, the CFB criteria can easily have excellent performance especially when the true model has less than $\frac{p}{2}$ nonzero components.

6. CFB can easily incorporate prior information. For example, if one would like to specify different prior probabilities ω_i on X_i based on some expert knowledge, this would not bring any difficulty in CFB - we can just treat ω_i as known and only integrate k_γ out of the posterior. However, it would cause trouble for the FB criterion with restriction.

However, CFB might be criticized for its lack of perfect theoretical guidelines of choosing the knob function. This can be explored in future research. Also, it is interesting to study the robustness (or sensitivity) of model posterior by specifying different knob functions.

Chapter 6

Generalization and A Special Case

In this chapter, I describe how to generalize Bayesian variable selection to GLM with a noncanonical link function. Also I discuss the normal linear regression as a special case of GLM, on which previous research concentrated. I talk about the connections between the work developed here and that in George and Foster (2000).

6.1 Generalization to GLM with A Noncanonical Link Function

The basic GLM considered so far has assumed a canonical link function. There exist some noncanonical link functions that are widely used in practice, for example, square root $\sqrt{\mu}$, exponent $(\mu + c_1)^{c_2}$ (c_1 and c_2 known), complementary log-log $\log(\frac{\mu}{n-\mu})$ and probit $\Phi^{-1}(\frac{\mu}{n})$ (μ is the mean of y, n is the sample size), etc. So it is important to generalize all the work developed in previous chapters to GLM with a noncanonical link function.

Consider a GLM with a noncanonical link function $g(.)$. This function must be monotonic and differentiable. Instead of having $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} = b'^{-1}(\boldsymbol{\mu})$, we have

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

$$\boldsymbol{\mu} = b'(\boldsymbol{\theta})$$

Based on this, the parameter $\boldsymbol{\beta}$ is connected with the natural parameter $\boldsymbol{\theta}$ as follows.

$$\boldsymbol{\theta} = b'^{-1} \circ g^{-1}(\mathbf{X}\boldsymbol{\beta})$$

where $\circ$ denotes the product of two functions (as mappings). I prefer to use $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\beta})$ in most of the situations in this section. Now we have

$$p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) = \exp\{\frac{\mathbf{Y}^T \cdot \boldsymbol{\theta}(\boldsymbol{\beta}) - \mathbf{b}^T(\boldsymbol{\theta}(\boldsymbol{\beta})) \cdot \mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi) \cdot \mathbf{1}\} \tag{6.1}$$

Denote

$$W(\hat{\boldsymbol{\beta}}_\gamma) = -\left(\frac{\partial^2 \log p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma)}{\partial\boldsymbol{\beta}_\gamma \partial\boldsymbol{\beta}_\gamma^T}\right)^{-1}_{\boldsymbol{\beta}_\gamma = \hat{\boldsymbol{\beta}}_\gamma} \tag{6.2}$$

where $W(\hat{\boldsymbol{\beta}}_\gamma)$ is a $(q_\gamma + 1) \times (q_\gamma + 1)$ matrix. I adopt the following prior distribution on $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$:

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\gamma, c \sim \mathbf{N}_{q_\gamma+1}(\mathbf{m}_\gamma, c\,W(\hat{\boldsymbol{\beta}}_\gamma)) \text{ for } c \in (0, +\infty) \tag{6.3}$$

Compared to the one given in (2.2), (6.3) is more general in the prior covariance matrix. Under this prior on $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$, the marginal distribution of $\mathbf{Y}$ is given by

$$\begin{aligned} p\,(\mathbf{Y}|\gamma, c) &= \int_{\mathbf{R}^{q_\gamma+1}} p\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) p\,(\boldsymbol{\beta}_\gamma|\gamma, c)\, d\boldsymbol{\beta}_\gamma \\ &= (2\pi)^{-\frac{q_\gamma+1}{2}} \left| c \cdot W(\hat{\boldsymbol{\beta}}_\gamma)\right|^{-\frac{1}{2}} \int_{\mathbf{R}^{q_\gamma+1}} \exp\left\{\frac{\mathbf{Y}^T \cdot \boldsymbol{\theta}(\boldsymbol{\beta}) - \mathbf{b}^T(\boldsymbol{\theta}(\boldsymbol{\beta})) \cdot \mathbf{1}}{\phi}\right. \\ &\quad \left. + \mathbf{c}^T(\mathbf{Y}, \phi) \cdot \mathbf{1} - \frac{(\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)^T W^{-1}(\hat{\boldsymbol{\beta}}_\gamma)(\boldsymbol{\beta}_\gamma - \mathbf{m}_\gamma)}{2c}\right\} d\boldsymbol{\beta}_\gamma \end{aligned} \tag{6.4}$$

Like in the canonical link case, the prior covariance matrix of $\boldsymbol{\beta_\gamma}$ is proportional to minus the inverse of the Hessian of $\log p\left(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma\right)$ evaluated at $\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}$. So it is trivial to show that the Laplace approximation and the Integrated Laplace approximation to (6.4) both stay the same as (2.6) and (2.9) respectively, except that we replace $\frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{\phi}$ by $W^{-1}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}})$ and $\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma$ by $\boldsymbol{\theta}(\hat{\boldsymbol{\beta}})$. For example, the Integrated Laplace approximation (2.9) is changed to:

$$\tilde{p}\left(\mathbf{Y}|\gamma, c\right) = (1+c)^{-\frac{q_\gamma+1}{2}} \exp\left\{\frac{\mathbf{Y}^T \cdot \boldsymbol{\theta}(\hat{\boldsymbol{\beta}}) - \mathbf{b}^T(\boldsymbol{\theta}(\hat{\boldsymbol{\beta}})) \cdot \mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi) \cdot \mathbf{1}\right\}$$
$$\exp\left\{-\frac{1}{2(c+1)}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T W^{-1}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}})(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\} \tag{6.5}$$

Now it is clear that a noncanonical link does not entail more difficulty in Bayesian variable selection at all. It is trial to show that all the EB, FB, CFB model posterior (or criteria) stay the same as before as long as we replace $\frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{\phi}$ by $W^{-1}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}})$ and $\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma$ by $\boldsymbol{\theta}(\hat{\boldsymbol{\beta}})$ in formulas listed in previous chapters.

6.2 A Special Case: Normal Linear Regression

The normal linear model is a special case of GLM with a canonical link, for which we have (see Table 1.1 and 1.2):

$$b(\theta_i) = \frac{1}{2}\theta_i^2,\ \phi = \sigma^2 \tag{6.6}$$

$$c(y_i, \phi) = -\frac{y_i^2}{2\phi} - \frac{1}{2}(\log\phi + \log 2\pi) = -\frac{y_i^2}{2\sigma^2} + \log\frac{1}{\sqrt{2\pi}\sigma} \tag{6.7}$$

where θ_i is the mean of y_i and σ^2 is the variance of y_i. Extensive research in variable selection has been done under this setting (see section 1.1 for a list of references). The reason why I consider it explicitly in this section are as follows.

1. Under the same Bayesian framework, George and Foster (2000) have done calibration and empirical Bayes variable selection, which motivated this work. As the formulas in George and Foster (2000) look different from the corresponding ones in this dissertation, their connections are not obvious. So it is of interest that I show those connections here.

2. As discussed in section 2.2.3, the Integrated Laplace approximation to $p(\mathbf{Y}|\gamma, c)$ is exactly itself for the normal linear model. As a result, any formulas developed in previous chapters will not be asymptotic; in a strict sense, they are all closed-form. Also, they should have simpler forms because of the nice properties of the normal linear model.

3. As it is most widely used in both practice and research, it would be convenient for others if I list all the results developed here separately for the normal linear regression.

Following George and Foster (2000), consider a normal linear model with n observations on a dependent variable Y and p potential independent variables $X = (x_1, x_2, ..., x_p)$:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$$

where $\epsilon \sim N(0, \sigma^2 \mathbf{I})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$. Given model γ, the MLE of $\boldsymbol{\beta_\gamma}$ is

$$\hat{\boldsymbol{\beta}}_\gamma = (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^T \mathbf{Y} \tag{6.8}$$

Note since the intercept term of model γ is fixed at zero under the above setting, $\boldsymbol{\beta_\gamma}$ is a $q_\gamma \times 1$ vector instead of a $(q_\gamma + 1) \times 1$ vector. Let $SS_\gamma = \hat{\boldsymbol{\beta}}_\gamma^T \mathbf{X}_\gamma^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma$. Plugging (6.8) into SS_γ , we have

$$\begin{aligned} SS_\gamma &= \hat{\boldsymbol{\beta}}_\gamma^T \mathbf{X}_\gamma^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma = \mathbf{Y}^T \mathbf{X}_\gamma [(\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1}]^T (\mathbf{X}_\gamma^T \mathbf{X}_\gamma) \cdot (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^T \mathbf{Y} \\ &= \mathbf{Y}^T \mathbf{X}_\gamma (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^T \mathbf{Y} \end{aligned} \tag{6.9}$$

Since $b''(\theta_i) = 1$ results in a $n \times n$ identity matrix of $\mathbf{V}_\gamma$ under the normal linear model, the prior distribution on $\boldsymbol{\beta_\gamma}$ given in (2.2) becomes

$$\boldsymbol{\beta_\gamma} | \gamma, c \sim \mathbf{N}_{q_\gamma}(\mathbf{m}_\gamma, c\,(\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1}) \text{ for } c \in (0, +\infty) \tag{6.10}$$

Let's look at the marginal distribution of $\mathbf{Y}$. Now we have a closed form for $p\,(\mathbf{Y}|\gamma, c)$. Applying (2.9), (6.8) and (6.9) together, we have

$$\begin{aligned} p\,(\mathbf{Y}|\gamma,\, c) &= (1+c)^{-\frac{q_\gamma}{2}} \exp\left\{ \frac{\mathbf{Y}^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma) \cdot 1}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi) \cdot 1 \right\} \\ &\quad \cdot \exp\left\{ -\frac{1}{2(c+1)\phi} (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma) \right\} \\ &= (1+c)^{-\frac{q_\gamma}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{\mathbf{Y}^T \mathbf{X}_\gamma (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma^T \mathbf{Y} - (\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma)^T (\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma)/2}{\sigma^2} \right. \\ &\quad \left. - \frac{\mathbf{Y}^T \mathbf{Y}}{2\sigma^2} \right\} \exp\left\{ -\frac{1}{2(c+1)\sigma^2} (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T \mathbf{X}_\gamma^T \mathbf{X}_\gamma (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma) \right\} \\ &= (1+c)^{-\frac{q_\gamma}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{SS_\gamma}{2\sigma^2} - \frac{\mathbf{Y}^T \mathbf{Y}}{2\sigma^2} \right\} \\ &\quad \exp\left\{ -\frac{1}{2(c+1)\sigma^2} (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T (\mathbf{X}_\gamma^T \mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma) \right\} \end{aligned} \tag{6.11}$$

In (6.11), if we take the prior mean $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}$ under the null linear model, i.e., $(0, 0, \cdots, 0)$, we have

$$p\left(\mathbf{Y}|\gamma,\, c\right) = (1+c)^{-\frac{q_\gamma}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\mathbf{Y}^T\mathbf{Y} - SS_\gamma}{2\sigma^2} - \frac{SS_\gamma}{2(c+1)\sigma^2}\right\} \tag{6.12}$$

So

$$\begin{aligned}&\pi\left(\mathbf{Y}|\gamma,\, c,\, \omega\right)\\ &\propto\ \omega^{q_\gamma}(1-\omega)^{p-q_\gamma}(1+c)^{-\frac{q_\gamma}{2}} \exp\left\{-\frac{\mathbf{Y}^T\mathbf{Y} - SS_\gamma}{2\sigma^2} - \frac{SS_\gamma}{2(c+1)\sigma^2}\right\}\end{aligned} \tag{6.13}$$

Finally, this is equation(7) in George and Foster (2000). If we take $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$, i.e., $\hat{\boldsymbol{\beta}}_\gamma$, we have

$$\begin{aligned}&\pi\left(\mathbf{Y}|\gamma,\, c,\, \omega\right)\\ &\propto\ \omega^{q_\gamma}(1-\omega)^{p-q_\gamma}(1+c)^{-\frac{q_\gamma}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{\mathbf{Y}^T\mathbf{Y} - SS_\gamma}{2\sigma^2}\right\}\end{aligned} \tag{6.14}$$

It is easy to see that AIC, BIC and RIC can be calibrated from either (6.13) or (6.14). By taking $\mathbf{m}_\gamma$ as $(0, 0, \cdots, 0)$, George and Foster (2000) did calibration based on (6.13). By taking $\mathbf{m}_\gamma$ as $\hat{\boldsymbol{\beta}}_\gamma$, I did calibration based on (6.13) in section 2.3. Both ways work for the normal linear model basically because the nice quadratic form of the log-likelihood function can be combined with the quadratic form of the logarithm of the prior density of $\boldsymbol{\beta}_\gamma$. For other members of GLM, taking $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}$ under the null linear model would not work for calibration.

As for the formulas for the EB, FB and CFB approaches given in previous chapters, we can achieve simplified forms by noticing the following facts for the normal linear model:

1. The estimated likelihood under model γ:
$$\hat{L}_\gamma = \exp\left\{\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{SS_\gamma}{2\sigma^2} - \frac{\mathbf{Y}^T\mathbf{Y}}{2\sigma^2}\right\} \qquad (6.15)$$

2. Given the data, the deviance D_γ of model γ is equivalent to $-2\log\hat{L}_\gamma$ in variable selection criteria, hence further it is equivalent to $-\frac{SS_\gamma}{\sigma^2}$ for the normal linear model. Also, T_γ, defined as $(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)/\phi$, is now simplified to $\frac{SS_\gamma}{\sigma^2}$.

3. The MLE of $\boldsymbol{\beta}$ under the null linear model becomes $(0,0,\cdots,0)$ for the linear setting specified at the beginning of the section, which is independent of the data. This eliminates $\mathbf{m}_\gamma$ from all the model posterior (or selection criteria) formulas if as usual, we take $\mathbf{m}_\gamma$ as the MLE of $\boldsymbol{\beta}$ under the null model.

Though tedious and a bit trivial, I re-list all the model posterior (or criteria) achieved from the EB, FB, CFB approaches for the normal linear model, which may be convenient for later researchers (I take $\mathbf{m}_\gamma$ as $(0,0,\cdots,0)$ here).

1. The CML criterion
$$C_{CML} = \begin{cases} -\frac{SS_\gamma}{\sigma^2} + q_\gamma\left(\log\frac{SS_\gamma}{q_\gamma\,\sigma^2}+1\right) - 2\left\{q_\gamma\log q_\gamma + (p-q_\gamma)\log(p-q_\gamma)\right\} & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} > 1 \\ -2\left\{q_\gamma\log q_\gamma + (p-q_\gamma)\log(p-q_\gamma)\right\} & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} \le 1 \end{cases} \qquad (6.16)$$

To minimize C_{CML} is equivalent to maximize equation (18) in George and Foster (2000).

2. The $MCML$ criterion

$$C_{MCML} = \begin{cases} -\frac{SS_\gamma}{\sigma^2} + q_\gamma(\log\frac{SS_\gamma}{q_\gamma\,\sigma^2} + 1) & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} > 1 \\ 0 & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} \leq 1 \end{cases} \tag{6.17}$$

3. The $HCML$ criterion

$$C_{HCML} = \begin{cases} -\frac{SS_\gamma}{\sigma^2} + q_\gamma(\log\frac{SS_\gamma}{q_\gamma\,\sigma^2} + 1) - 2\left\{\log\Gamma(q_\gamma+1) + \log\Gamma(p-q_\gamma+1)\right\} & \\ & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} > 1 \\ -2\left\{\log\Gamma(q_\gamma+1) + \log\Gamma(p-q_\gamma+1)\right\} & \\ & \text{if } \frac{SS_\gamma}{q_\gamma\,\sigma^2} \leq 1 \end{cases} \tag{6.18}$$

4. The FB model posterior under “noninformative” priors

$$\begin{aligned} \pi\left(\gamma|\mathbf{Y}\right) \propto{}& \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)} \cdot \Gamma(\frac{q_\gamma}{2}+1) \\ &\cdot \left(\frac{SS_\gamma}{2\sigma^2}\right)^{-(\frac{q_\gamma}{2}+1)} \cdot G_{\alpha=\frac{q_\gamma}{2}+1,\beta=1}\left(\frac{SS_\gamma}{2\sigma^2}\right) \cdot \exp\left\{\frac{SS_\gamma}{2\sigma^2}\right\} \end{aligned} \tag{6.19}$$

5. The FB model posterior under “noninformative” priors with restricting the integration region

$$\begin{aligned} \pi\left(\gamma|\mathbf{Y}\right) \propto{}& \Gamma(\frac{q_\gamma}{2}+1) \cdot \left(\frac{SS_\gamma}{2\sigma^2}\right)^{-(\frac{q_\gamma}{2}+1)} \cdot \exp\left\{\frac{SS_\gamma}{2\sigma^2}\right\} \\ &\left\{\frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)} B_{(q_\gamma+1,p-q_\gamma+1)}(\frac{1}{2}) \cdot G_{(\frac{q_\gamma}{2}+1,1)}\left(\frac{SS_\gamma}{2\sigma^2}\right)\right. \\ &\left. + \int_{0.5}^{1} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} \cdot G_{(\frac{q_\gamma}{2}+1,1)}\left((\frac{SS_\gamma}{2\sigma^2})(\frac{1}{\omega}-1)^2\right)\right\} d\omega \end{aligned} \tag{6.20}$$

6. The FB model posterior under conjugate priors with(without) restriction of integration region

   I ignore these here.

7. The CFB model posterior under "noninformative" priors

$$\pi\left(\gamma|\mathbf{Y}\right) \propto \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)} \cdot \tau \cdot \Gamma(\frac{q_\gamma}{2}+\tau) \cdot \left(\frac{SS_\gamma}{2\sigma^2}\right)^{-(\frac{q_\gamma}{2}+\tau)} \cdot G_{\alpha=\frac{q_\gamma}{2}+\tau,\beta=1}\left(\frac{SS_\gamma}{2\sigma^2}\right) \cdot \exp\left\{\frac{SS_\gamma}{2\sigma^2}\right\} \quad (6.21)$$

Chapter 7

Simulation

In this chapter I evaluate and report the performance potential of selection using the EB criteria proposed in Chapter 3 that include C_{CML}, C_{MCML}, C_{HCML}, some FB criteria proposed in Chapter 4 that include those under noninformative uniform hyperpriors with/without restriction of the integration region, and the CFB criterion proposed in Chapter 5 with the knob function being $\tau_1 = \frac{1}{q_\gamma+1}$. Here I do not focus on the criteria under the conjugate hyperpriors in my simulation since in practice, people usually do not have strong prior information of hyperparameters. For comparison, I include the two fixed penalty selection criteria AIC and BIC. I also include "no-selection" and "gold" criteria, to see whether other criteria have practical value in selecting the "best" model and how well they work compared to the gold standard. The "no-selection" criterion always chooses the full model as the best model and the "gold" criterion always picks up the true model.

For concision, I use CFB for the CFB criterion with the knob function being τ_1; I use $FB.r$ and FB for the FB criteria under noninformative hyperpriors with/without restricting integration area, respectively; I use NS for the 'no-selection' criterion. Others are self-explanatory, like CML, $MCML$,

$HCML$, $GOLD$, etc.

7.1 Simulation Set-ups

I follow the simulation set-ups for normal linear models in George and Foster (2000) closely but with some tunes to GLM. Aspects of the set-ups are as follows.

1. Types of GLM

   I consider Poisson model, Logistic regression and linear regression, all with canonical links in this simulation. So $\mathbf{Y}$ is simulated from Poisson, Bernoulli and Gaussian distribution, respectively.

2. Number of observations n and number of potential independent variables p

   I consider the sample size n of 100, 200 and 500. For n of 100 and 500, I consider p of 10, 20; for n of 200, I consider p of 10, 20, 50.

3. Generating $\mathbf{X}$

   Once n and p are fixed, $\mathbf{X}$ can be generated. Considering we always can standardize each X_i before fitting the GLM model, without loss of generality the n rows of $\mathbf{X}$ are independently generated from a $N_p(0, \Sigma)$ distribution with the ijth element of Σ be $\rho^{|i-j|}$, that is

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{p-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{p-2} \\ \vdots & & \vdots & & & \vdots \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} & \rho^{p-4} & \cdots & 1 \end{pmatrix}$$

Thus, by choosing non-zero ρ, correlations can be built into predictor variables. I consider 0 and 0.5 as values of ρ. For each combination of (n, p, ρ) under a distribution of Y, $\mathbf{X}$ is held fixed.

4. Generating $\boldsymbol{\beta}$

   For each fixed $\mathbf{X}$, $u+1$ different vectors $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)$ will be generated as: $\boldsymbol{\beta}_0 = (\beta_0^0, 0, \cdots, 0)$, $\boldsymbol{\beta}_i = (\beta_0^i, \mathbf{B}_i, \mathbf{B}_i, \cdots, \mathbf{B}_i, \mathbf{B}_i)$ (there are in total v replicates of $\mathbf{B}_i$ in $\boldsymbol{\beta}_i$) , $i = 1, 2, \cdots, u$, where u, v satisfy that $u \cdot v = p$, $\mathbf{B}_i = (b_1^i, b_2^i, \cdots, b_u^i)$ and $\mathbf{B}_i$ consists of i adjacent nonzero values of b^i centered around $b^i_{\lfloor \frac{u+1}{2} \rfloor}$ and zero values of b^i otherwise. Then for each i, $i = 1, 2, \cdots, u$, β_0^i and the i nonzero values of b^i are generated from a normal distribution $N(0, \sigma)$ (usually σ takes a value of 0.25 or 0.5, or 1). By doing so, for each fixed p, the null model, model with v nonzero components, model with $2v$ nonzero components, ..., model with uv nonzero components (the full model) are considered in my simulation. For p being 10 and 20, I take u as 5 and for p being 50, I take u as 10. For example, for p being 50, 10 different $\mathbf{B}_i, i = 1, 2, \cdots, 10$ are constructed: $\mathbf{B}_1 = (0, 0, 0, 0, b_5^1, 0, 0, 0, 0, 0)$, $\mathbf{B}_2 = (0, 0, 0, 0, b_5^2, b_6^2, 0, 0, 0, 0)$, $\mathbf{B}_3 = (0, 0, 0, b_4^3, b_5^3, b_6^3, 0, 0, 0, 0)$, $\mathbf{B}_4 = (0, 0, 0, b_4^4, b_5^4, b_6^4, b_7^4, 0, 0, 0)$, ..., $\mathbf{B}_{10} = (b_1^{10}, b_2^{10}, b_3^{10}, b_4^{10}, b_5^{10}, b_6^{10}, b_7^{10}, b_8^{10}, b_9^{10}, b_{10}^{10})$. Then, 11 different $\boldsymbol{\beta}_i, i = 0, 1, \cdots, 10$ with length of 51 are generated as: $\boldsymbol{\beta}_0 = (\beta_0, 0, \cdots, 0)$, $\boldsymbol{\beta}_i = (\beta_0, \mathbf{B}_i, \mathbf{B}_i, \mathbf{B}_i, \mathbf{B}_i, \mathbf{B}_i), i = 1, 2, \cdots, 10$.

   For each i, $i = 0, 1, \cdots, u$, it is required the values of $\boldsymbol{\beta}_i$ satisfy theoretical Pseudo R^2 is moderate , where Pseudo $R^2 = 1 - \frac{\log L_T}{\log L_N}$, L_T is the likelihood of the true model and L_N is the likelihood of the null model. Based on this, I require that the following equation holds approximately when generating $\boldsymbol{\beta}$:

$$1-\frac{\frac{\boldsymbol{\mu}^T\mathbf{X}\boldsymbol{\beta}-\mathbf{b}^T(\mathbf{X}\boldsymbol{\beta})\cdot\mathbf{1}}{\phi}+\mathbf{c}^T(\boldsymbol{\mu},\phi)\cdot\mathbf{1}}{\frac{nb'^{-1}(\bar{\mu})\cdot\mu-nb(b'^{-1}(\bar{\mu}))}{\phi}+\mathbf{c}^T(\boldsymbol{\mu},\phi)\cdot\mathbf{1}}=S \tag{7.1}$$

where $\boldsymbol{\mu}=\mathbf{b}^{'}(\boldsymbol{\theta})$ is the mean vector of $\mathbf{Y}$ and $\bar{\mu}=\frac{\boldsymbol{\mu}\cdot\mathbf{1}}{n}$. Usually, S takes a value between 0.4 and 0.6. This is to assure that the relationship between $\mathbf{Y}$ and $\mathbf{X}$ is such that fitting a regression model is not too easy or too difficult. If S is set too high for a model, the signal contained in the generated data is so strong that every criterion works well and select the right model at most times. Also, this is away from real data. In the other direction, if S is set too low, all the criteria work poorly in selecting right models though they still reduce predictive loss. Hence a moderate Pseudo R^2 for a model is chosen in the simulation.

For each combination of (n, p, ρ, q), where q is number of the non-zero components of the true model, $\boldsymbol{\beta}$ is held fixed.

5. Generating $\mathbf{Y}$

   After $\mathbf{X}$ and $\boldsymbol{\beta}$ are generated, $\mathbf{Y}$ is simulated from the exponential family distribution under consideration. For each combination of (n, p, ρ, q), $\mathbf{Y}$ is regenerated 250 times based on fixed $\mathbf{X}$ and $\boldsymbol{\beta}$, then the performance of each criterion is averaged based on the 250 iterations.

6. Searching the model space

   Under the GLM setting, for a medium or large p, it is not feasible to run the simulation by searching the whole space with 2^p models for the 'best' model. Even for $p = 10$ and $n = 200$, it takes a few weeks to search the whole 1024 models under the above settings on a PC (with 1.33 GHZ CPU and 320 MB memory) and sometimes there are serious convergence problems that prevent us from getting a stable evaluation of different criteria. Hence, I consider applying criteria to a subset of models obtained by a stepwise method. For each simulated Y, I simply use each criterion to select a model from the subset visited by forward selection. Since it is possible that the true model will not be visited by forward selection, I always force the true model into the subset(I discuss this in next section). So the number of models in the subset is either $p+1$, p plus the null model or $p+2$, p plus the null model plus the true model.

7.2 Assessment of Performance

It is very important to give a fair evaluation of all criteria developed here. However, it is not easy to accomplish that through a single measure.

First, I follow George and Foster (2000) and adopt predictive loss as a "rule" that measures how similar a model is to the true model with known coefficients. At each iteration within which $\mathbf{Y}$ is re-generated, and for each selection criterion, the disparity between the selected model $\hat{\gamma}$ and the correct

γ is summarized by the predictive loss. It can be defined in a linear scale:

$$L\left\{\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}(\hat{\gamma})\right\} \equiv \left\{\mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\gamma}) - \mathbf{X}\boldsymbol{\beta}\right\}^T \left\{\mathbf{X}\hat{\boldsymbol{\beta}}(\hat{\gamma}) - \mathbf{X}\boldsymbol{\beta}\right\}$$

or in a fitted scale:

$$L\left\{\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}(\hat{\gamma})\right\} \equiv (\hat{\boldsymbol{\mu}}(\hat{\gamma}) - \boldsymbol{\mu})^T(\hat{\boldsymbol{\mu}}(\hat{\gamma}) - \boldsymbol{\mu})$$

For Poisson models, I define the predictive loss in the linear scale. For Logistic regression, I define it in the fitted scale. The purpose of doing so is to keep variances of the predictive loss for different q at the same level and also as small as possible. For linear regressions, the two definitions are equivalent.

The predictive loss provides a natural scalar summary of the disparity between $\hat{\gamma}$ and γ. However, the model with the minimum predictive loss is not necessarily the true model with estimated coefficients. It is not rare to find that some models have a smaller predictive loss than the true one (without explicitly stating it, I always mean the fitted true model when I say the true model). Therefore, I also report the simulation results by looking at whether the true model is selected by a criterion given data. During an iteration, each criterion is applied to a subset chosen from the stepwise method, which might not include the true model. So whether the criterion hits the true model is not meaningful if the subset does not contain the true one. One compromise is to look at whether a criterion selects a model with the minimum predictive loss in the subset. Actually this turns out not to be a wise choice because a model with the minimum predictive loss is sometimes not the true model even if the true model is included in the subset. Another natural compromise is to force the true model into a subset when it is not included. This might be a little away from real practice, but I assume that a criterion that often selects a right model will tend to select a model closest to the true one when the true one is not included in the subset. I think this assumption is more reasonable than the one that assumes that a model with the minimum predictive loss is closest to the true one. And that is the reason why I choose the second method and always include the true model in a subset.

The drawback of the measure whether the true model is selected given data is obvious: when either p or q is large, all the criteria tend to work poorly in finding the true model. But certainly some criteria would select models that have satisfactory prediction performance. So looking at both the predictive loss and whether the true model is selected is a good way to evaluate the performance of variable selection.

Under each (d, n, p, ρ) where d stands for a given distribution of Y, I report the following for a criterion considered here:

1. the average predictive loss and its standard error.
2. the percent of hitting the true model and its standard error.
3. mean differences in predictive loss between some pairs of criteria and their significance levels, which we might be especially interested in, for example, FB and $FB.r$. The two statistics are from paired T-tests.

I also look at the relative performance of different criteria at different q for a typical (n, p) under each (d, ρ) . The average predictive loss and the percent of hits for a criterion along with their standard errors under each (d, n, p, ρ, q) are given in Appendix C.

It is worth mentioning that the effect of q or its interaction with criteria on variable selection can not be fairly evaluated under this setting of the simulation. The reason is that for all the 250 iterations of a q under a (d, n, p, ρ), only one $\mathbf{X}$ and $\boldsymbol{\beta}$ is generated, which means all the 250 generated models with the same q have the same $\mathbf{X}$ and $\boldsymbol{\beta}$. Some $(\mathbf{X}, \boldsymbol{\beta})$ might be related to high predictive losses. For example, for Poisson models, $\exp(\mathbf{X}\boldsymbol{\beta})$ determines the mean and variance of $\mathbf{Y}$, so a large $\mathbf{X}\boldsymbol{\beta}$ tends to have large predictive losses. Obviously, the effect of q or its interaction can not be separated from the effect of $\mathbf{X}$ and $\boldsymbol{\beta}$. Here, I make a reasonable assumption that the effect of q has been randomized as models are generated across different q.

To compare any two criteria under a (d, n, p, ρ), I use the paired T-tests since given any simulated data, their values are not independent.

Based on simulation results, I address the following research questions.

1. Do the FB criteria(CFB, $FB.r$ and FB) outperform the EB criteria(CML, $MCML$, $HCML$)?

2. How well do the EB and FB criteria work, compared to the classical criteria, AIC and BIC?

3. Does restriction of integration area improve performance of the FB methods?

4. Does the CFB method improve performance of the FB methods?

5. Among the EB criteria, does $HCML$ or $MCML$ improve performance of CML?

6. How well do the EB and FB criteria work, compared to the nonselection method and the gold standard?

7.3 Poisson Models

I report simulation results based on Poisson models in this section. For each Y_i, under the Poisson model we have :

$$f(y_i|\mu_i) = \exp(-\mu_i) \cdot \frac{\mu_i^{y_i}}{y_i!}$$

where μ_i is the mean of Y_i and y_i is a nonnegative integer. From Table (1.1), we have $\phi = 1$, $b(\theta_i) = \mu_i = \exp(\theta_i)$ and $c(y_i, \phi) = -\log(y_i!)$. Also, under the canonical link function, we have $\boldsymbol{\mu} = \exp(\mathbf{X}\boldsymbol{\beta})$. I define the predictive loss in the linear scale since the one in the fitted scale is related to $\exp(\mathbf{X}\hat{\boldsymbol{\beta}})$ and has very large variation. I deliberately generate Y_i from small μ_i so that we can observe any difference in criteria performance between Poisson and normal linear models.

For each (n, p) and each criterion, I average the predictive losses over different models in Table (7.1) for $\rho = 0$ and Table (7.3) for $\rho = 0.5$. Along with each average loss, its standard error is reported in italic. The percent of hits for $\rho = 0$ and $\rho = 0.5$ is given in Table (7.2) and Table (7.4), respectively (the italic numbers are standard errors). For better visualization, I plot the average predictive losses of all criteria along with their 95% confidence intervals within a figure for each (n, p, ρ) (see Figure (7.1) for $\rho = 0$ and Figure (7.2) for $\rho = 0.5$). For a conclusive comparison, I report mean differences in predictive loss with their significance levels from paired T-test for any interested pair of criteria (see Table (7.5) for $\rho = 0$ and Table (7.6) for $\rho = 0.5$; "***" means the p-value is smaller than 0.001, "**" means the p-value is smaller than 0.01 but at least 0.001, "*" means the p-value is smaller than 0.05 but at least 0.01 and no star means the p-value is at least 0.05).

For readers to get a basic idea of how different selection criteria work at different numbers of nonzero components, Table (7.7) for $\rho = 0$ and Table (7.9) for $\rho = 0.5$ give the average predictive loss by different q, and Table (7.8) for $\rho = 0$ and Table (7.10) for $\rho = 0.5$ list number of hits by different q under the case $n = 200, p = 20$ (the total number is 250 for each q). Also, I present Figure (7.3) for $\rho = 0$ and Figure (7.4) for $\rho = 0.5$ to compare criteria for $n = 200, p = 20$. Within each figure, there are two subfigures: (a) plots the average predictive losses; (b) plots the frequencies of choosing the "true" model. For simulation results by number of nonzero components for all the (d, n, p, ρ), please refer to Appendix C.1.

My findings in overall performance based on predictive loss are as follows.

1. The three FB criteria are significantly better than the three EB criteria.

2. All the three FB criteria perform significantly better than AIC. CFB and $FB.r$ are much better than BIC. FB does better than BIC in most cases. CFB and $FB.r$ improve FB notably. BIC works uniformly better than AIC.

3. Among the three EB criteria, CML and $MCML$ are worse than AIC. $HCML$ improves CML and $MCML$ in the sense that it performs better than AIC when $\rho = 0$, and sometimes better and sometimes worse than AIC when $\rho = 0.5$. However, $HCML$ is still not as good as BIC.

4. Not surprisingly, all the selection criteria are better than the nonselection method and worse than the gold standard. CFB and $FB.r$ work better than any other criteria and their average predictive losses are far below NS. For some (n, p), they are very close to the gold standard. This indicates the important practical value of variable selection in prediction.

My findings in overall performance based on percent of hits are listed below.

1. The FB criteria work much better than the EB criteria.

2. The FB criteria perform significantly better than AIC, and are better than BIC. Unlike in predictive loss, CFB and $FB.r$ don't outperform FB notably in terms of percent of hits. I would say they do similarly. The reason is that FB tends to choose the full model so it works better

than the other two when the true model is full, which brings up its percent of hits.

3. Generally speaking, the three EB criteria perform somewhere between AIC and BIC. BIC is much better than AIC. $HCML$ outperforms AIC considerably and sometimes is even better than BIC. $MCML$ outperforms AIC but is worse than BIC. CML works better than AIC for $\rho = 0$,but is sometimes worse than AIC for $\rho = 0.5$. $HCML$ significantly improves CML. $MCML$ improves CML or at least is comparable to CML except when $p = 50$.

By looking at the relative performance of criteria at different q, I find that CFB and $FB.r$ are usually better than or at least comparable to others at any number of nonzero components.

For Poisson models, I conclude that the FB criteria are consistently better than the EB or traditional criteria. Among the three FB criteria, CFB and $FB.r$ significantly improve FB in predictive loss. The EB criteria usually do not exceed BIC in variable selection. In addition, CFB and $FB.r$ usually achieve the top performance at any number of nonzero components, which is a very nice property in variable selection practice.

Table 7.1: Average Predictive Loss: Poisson, $\rho = 0$

p	n	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
10	100	3.24	4.02	4.07	4.13	4.54	4.48	4.83	4.58	4.95	5.59
		0.09	*0.11*	*0.11*	*0.11*	*0.11*	*0.11*	*0.10*	*0.11*	*0.10*	*0.10*
10	200	3.01	3.37	3.37	3.63	3.84	4.28	4.67	4.40	4.68	5.87
		0.09	*0.11*	*0.10*	*0.11*	*0.11*	*0.11*	*0.10*	*0.11*	*0.11*	*0.11*
10	500	4.16	4.36	4.39	4.83	4.68	5.58	5.96	7.79	5.84	7.31
		0.12	*0.12*	*0.12*	*0.13*	*0.13*	*0.13*	*0.12*	*0.20*	*0.13*	*0.12*
20	100	12.24	11.06	11.29	12.17	13.92	16.76	17.95	16.54	18.76	25.65
		0.38	*0.36*	*0.37*	*0.40*	*0.39*	*0.45*	*0.37*	*0.44*	*0.46*	*0.45*
20	200	6.21	6.71	6.74	7.32	7.48	9.89	10.67	9.99	10.75	14.59
		0.13	*0.15*	*0.15*	*0.18*	*0.15*	*0.22*	*0.16*	*0.21*	*0.21*	*0.21*
20	500	7.60	7.70	7.73	8.22	8.54	10.81	11.20	13.60	11.99	14.35
		0.21	*0.21*	*0.21*	*0.23*	*0.25*	*0.27*	*0.23*	*0.35*	*0.28*	*0.27*
50	200	15.06	21.46	21.11	22.21	25.09	27.73	27.67	27.26	28.35	37.90
		0.24	*0.32*	*0.32*	*0.34*	*0.37*	*0.41*	*0.27*	*0.40*	*0.38*	*0.34*

Table 7.2: Percent of Hits: Poisson, $\rho = 0$

p	n	cfb	fb.r	fb	bic	hcml	aic	cml	mcml
10	100	0.40	0.37	0.41	0.27	0.37	0.14	0.17	0.16
		0.013	*0.012*	*0.013*	*0.011*	*0.012*	*0.009*	*0.010*	*0.009*
10	200	0.57	0.54	0.55	0.44	0.51	0.20	0.31	0.31
		0.013	*0.013*	*0.013*	*0.013*	*0.013*	*0.010*	*0.012*	*0.012*
10	500	0.94	0.94	0.91	0.92	0.79	0.51	0.56	0.77
		0.006	*0.006*	*0.007*	*0.007*	*0.011*	*0.013*	*0.013*	*0.011*
20	100	0.66	0.53	0.64	0.41	0.44	0.10	0.25	0.26
		0.012	*0.013*	*0.012*	*0.013*	*0.013*	*0.008*	*0.011*	*0.011*
20	200	0.43	0.43	0.42	0.38	0.30	0.08	0.20	0.23
		0.013	*0.013*	*0.013*	*0.013*	*0.012*	*0.007*	*0.010*	*0.011*
20	500	0.73	0.73	0.71	0.69	0.53	0.28	0.31	0.55
		0.011	*0.011*	*0.012*	*0.012*	*0.013*	*0.012*	*0.012*	*0.013*
50	200	0.42	0.34	0.40	0.20	0.29	0.01	0.19	0.10
		0.009	*0.009*	*0.009*	*0.008*	*0.009*	*0.002*	*0.008*	*0.006*

Table 7.3: Average Predictive Loss: Poisson, $\rho = 0.5$

p	n	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
10	100	4.51	6.40	6.30	7.07	7.13	8.78	8.01	9.12	10.07	9.82
		0.20	*0.20*	*0.22*	*0.26*	*0.25*	*0.29*	*0.25*	*0.29*	*0.31*	*0.29*
10	200	3.59	4.22	4.21	4.47	4.40	5.06	5.17	6.26	5.18	6.27
		0.10	*0.11*	*0.12*	*0.12*	*0.12*	*0.12*	*0.11*	*0.15*	*0.11*	*0.11*
10	500	3.53	3.24	3.31	3.95	3.77	5.25	5.65	5.78	5.80	7.55
		0.09	*0.09*	*0.10*	*0.14*	*0.10*	*0.17*	*0.13*	*0.16*	*0.17*	*0.16*
20	100	6.04	9.58	9.20	10.93	9.70	14.76	11.68	14.72	14.50	15.64
		0.16	*0.26*	*0.24*	*0.31*	*0.22*	*0.33*	*0.23*	*0.33*	*0.31*	*0.31*
20	200	4.58	5.51	5.62	6.66	7.10	13.62	11.57	14.41	16.34	17.36
		0.11	*0.14*	*0.15*	*0.32*	*0.22*	*0.66*	*0.42*	*0.66*	*0.68*	*0.68*
20	500	5.03	6.09	6.07	6.54	6.88	8.59	8.50	10.03	9.34	10.88
		0.13	*0.17*	*0.16*	*0.18*	*0.18*	*0.20*	*0.15*	*0.21*	*0.20*	*0.17*
50	200	19.53	25.18	25.72	25.44	27.54	31.10	31.89	30.27	30.79	44.00
		0.34	*0.42*	*0.43*	*0.41*	*0.45*	*0.43*	*0.33*	*0.42*	*0.39*	*0.31*

Table 7.4: Percent of Hits: Poisson, $\rho = 0.5$

p	n	cfb	fb.r	fb	bic	hcml	aic	cml	mcml
10	100	0.49	0.44	0.46	0.45	0.32	0.28	0.11	0.27
		0.013	*0.013*	*0.013*	*0.013*	*0.012*	*0.012*	*0.008*	*0.012*
10	200	0.83	0.84	0.80	0.81	0.64	0.49	0.44	0.67
		0.010	*0.009*	*0.010*	*0.010*	*0.012*	*0.013*	*0.013*	*0.012*
10	500	0.63	0.61	0.60	0.54	0.49	0.35	0.26	0.41
		0.012	*0.013*	*0.013*	*0.013*	*0.013*	*0.012*	*0.011*	*0.013*
20	100	0.52	0.54	0.47	0.47	0.30	0.15	0.12	0.25
		0.013	*0.013*	*0.013*	*0.013*	*0.012*	*0.009*	*0.008*	*0.011*
20	200	0.60	0.58	0.59	0.51	0.54	0.10	0.36	0.34
		0.013	*0.013*	*0.013*	*0.013*	*0.013*	*0.008*	*0.012*	*0.012*
20	500	0.72	0.72	0.70	0.63	0.58	0.25	0.37	0.51
		0.012	*0.012*	*0.012*	*0.012*	*0.013*	*0.011*	*0.012*	*0.013*
50	200	0.26	0.25	0.25	0.12	0.23	0.00	0.15	0.05
		0.008	*0.008*	*0.008*	*0.006*	*0.008*	*0.000*	*0.007*	*0.004*

Table 7.5: Paired Comparison in Predictive Loss: Poisson, $\rho = 0$

type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
cfb-fb.r	-0.05	-0.00	-0.03	-0.23**	-0.02	-0.03	0.35***
fb.r-fb	-0.06	-0.26***	-0.44***	-0.88***	-0.58***	-0.48***	-1.09***
cfb-fb	-0.11**	-0.26***	-0.47***	-1.10***	-0.61***	-0.52***	-0.75***
cfb-bic	-0.51***	-0.46***	-0.32***	-2.86***	-0.76***	-0.84***	-3.63***
fb.r-bic	-0.46***	-0.46***	-0.28***	-2.63***	-0.74***	-0.81***	-3.98***
fb-bic	-0.40***	-0.20***	0.16*	-1.75***	-0.15	-0.33*	-2.88***
bic-aic	-0.29***	-0.83***	-1.29***	-4.03***	-3.19***	-2.66***	-2.57***
hcml - cml	-0.10***	-0.12***	-2.21***	0.23*	-0.10	-2.80***	0.47***

Table 7.6: Paired Comparison in Predictive Loss: Poisson, $\rho = 0.5$

type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
cfb-fb.r	0.09	0.01	-0.07*	0.38***	-0.11	0.01	-0.53***
fb.r-fb	-0.77***	-0.26***	-0.64***	-1.73***	-1.04***	-0.47***	0.27
cfb-fb	-0.68***	-0.25***	-0.71***	-1.35***	-1.14***	-0.46***	-0.26
cfb-bic	-0.73***	-0.18***	-0.53***	-0.11	-1.59***	-0.79***	-2.36***
fb.r-bic	-0.83***	-0.19***	-0.46***	-0.49***	-1.49***	-0.80***	-1.83***
fb-bic	-0.06	0.07	0.18	1.24***	-0.45	-0.33**	-2.10***
bic-aic	-0.88***	-0.77***	-1.88***	-1.98***	-4.46***	-1.62***	-4.35***
hcml - cml	-0.34***	-1.20***	-0.53***	0.04	-0.79***	-1.44***	0.84***

Table 7.7: Average Predictive Loss by Nonzero Components
Poisson, $\rho = 0, n = 200, p = 20$

q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
0	1.12	1.12	1.28	4.44	4.05	17.77	13.28	18.72	22.42	22.84
	0.08	*0.08*	*0.14*	*0.68*	*0.35*	*0.85*	*0.52*	*0.77*	*0.55*	*0.52*
4	2.40	2.80	2.80	2.80	3.42	2.93	6.90	2.92	3.80	11.35
	0.13	*0.17*	*0.17*	*0.17*	*0.20*	*0.18*	*0.26*	*0.18*	*0.22*	*0.29*
8	7.09	4.81	4.84	4.84	5.98	5.14	12.18	5.08	6.77	20.23
	0.31	*0.33*	*0.33*	*0.33*	*0.39*	*0.35*	*0.52*	*0.35*	*0.42*	*0.62*
12	7.04	8.41	8.44	8.70	8.10	11.03	9.24	10.52	8.24	11.23
	0.24	*0.30*	*0.30*	*0.32*	*0.31*	*0.33*	*0.30*	*0.34*	*0.28*	*0.32*
16	7.40	10.44	10.35	10.34	10.72	9.69	9.90	9.80	10.48	9.65
	0.21	*0.26*	*0.25*	*0.26*	*0.25*	*0.26*	*0.26*	*0.27*	*0.26*	*0.26*
20	12.21	12.70	12.71	12.81	12.59	12.79	12.53	12.92	12.77	12.21
	0.34	*0.24*	*0.24*	*0.25*	*0.24*	*0.29*	*0.29*	*0.29*	*0.24*	*0.34*

Table 7.8: No. of Hits by Nonzero Components
Poisson, $\rho = 0, n = 200, p = 20$

q	cfb	fb.r	fb	bic	hcml	aic	cml	mcml
0	250	248	227	174	82	9	0	1
4	229	229	229	190	222	22	223	167
8	0	0	0	0	0	0	0	0
12	167	160	156	200	43	73	81	163
16	5	9	2	1	0	14	0	8
20	1	0	9	0	97	0	0	0

Table 7.9: Average Predictive Loss by Nonzero Components
Poisson, $\rho = 0.5, n = 200, p = 20$

q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
0	2.73	2.73	3.54	9.73	11.05	52.00	36.73	56.83	66.93	68.05
	0.27	*0.27*	*0.46*	*1.78*	*1.01*	*2.86*	*1.68*	*2.59*	*2.02*	*1.97*
4	1.19	1.25	1.28	1.28	1.57	1.35	2.89	1.32	1.59	4.31
	0.07	*0.07*	*0.07*	*0.07*	*0.08*	*0.07*	*0.10*	*0.07*	*0.08*	*0.10*
8	3.16	3.59	3.67	3.67	3.78	4.03	5.44	3.97	4.14	7.53
	0.13	*0.16*	*0.16*	*0.16*	*0.16*	*0.18*	*0.20*	*0.17*	*0.17*	*0.22*
12	3.68	4.50	4.50	4.59	4.24	5.55	5.06	5.27	4.45	6.19
	0.11	*0.15*	*0.15*	*0.15*	*0.15*	*0.16*	*0.14*	*0.15*	*0.14*	*0.14*
16	5.52	7.14	7.08	7.19	7.22	6.95	7.07	7.02	7.19	6.89
	0.17	*0.18*	*0.18*	*0.19*	*0.17*	*0.19*	*0.18*	*0.19*	*0.18*	*0.19*
20	11.17	13.88	13.64	13.49	14.76	11.84	12.23	12.04	13.77	11.17
	0.29	*0.31*	*0.30*	*0.30*	*0.29*	*0.30*	*0.29*	*0.29*	*0.29*	*0.29*

Table 7.10: No. of Hits by Nonzero Components
Poisson, $\rho = 0.5, n = 200, p = 20$

q	cfb	fb.r	fb	bic	hcml	aic	cml	mcml
0	250	245	233	173	91	7	0	0
4	244	240	240	187	229	18	233	180
8	218	212	212	195	181	42	184	157
12	168	165	163	203	103	69	116	167
16	3	6	4	0	0	9	0	2
20	20	0	31	0	202	0	0	0

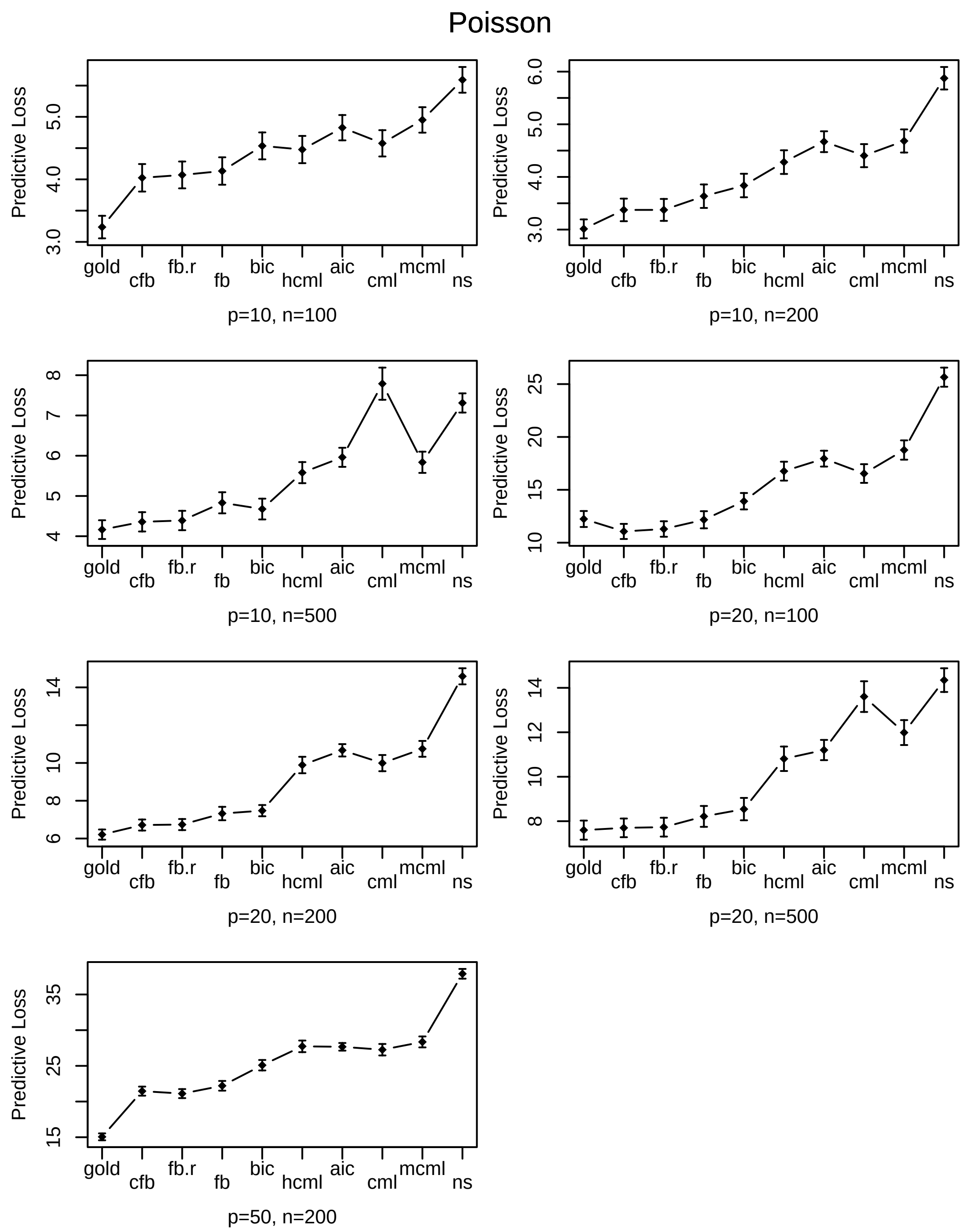

Figure 7.1: Average Predictive Loss With 95% CI: Poisson, $\rho = 0$

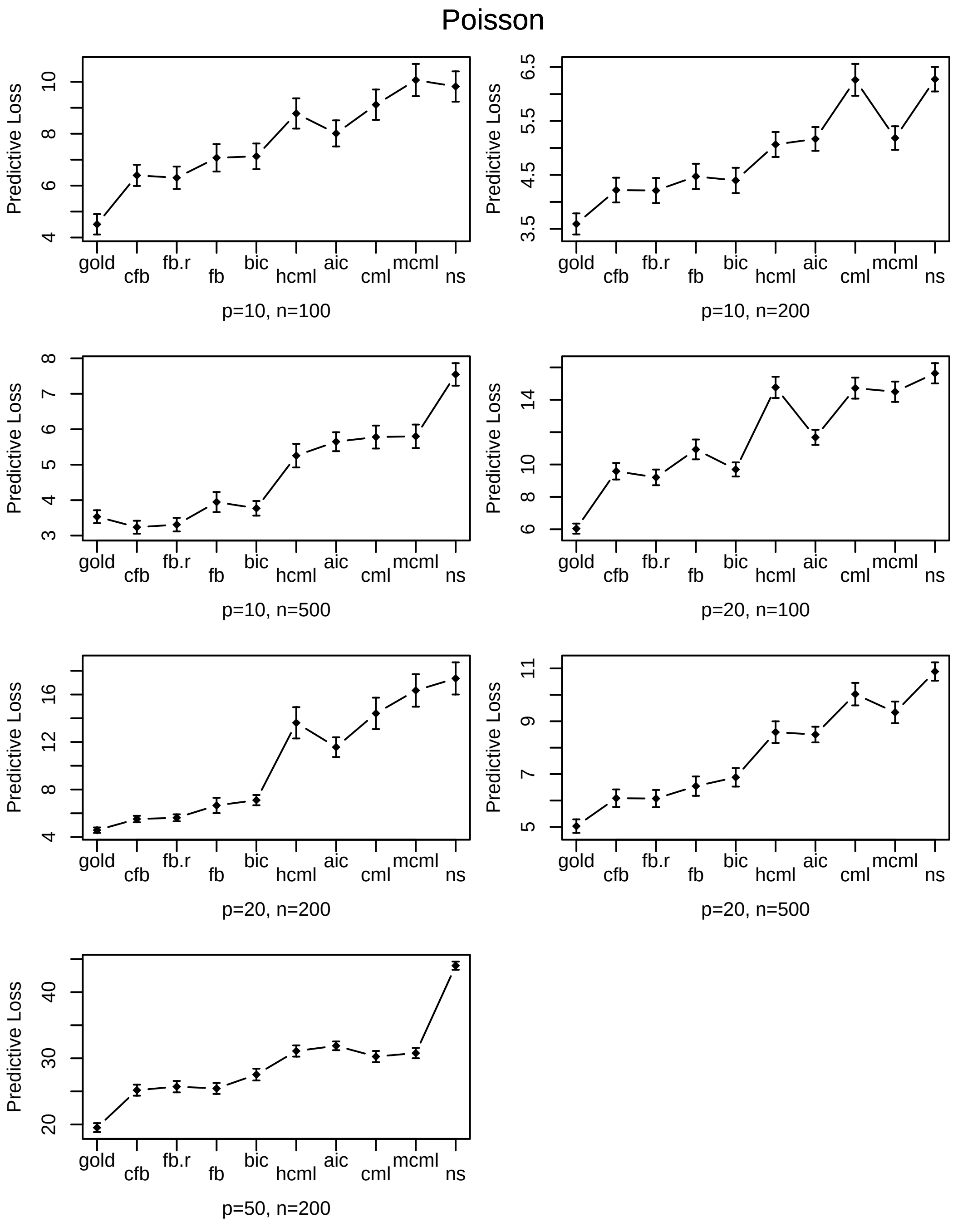

Figure 7.2: Average Predictive Loss With 95% CI: Poisson, $\rho = 0.5$

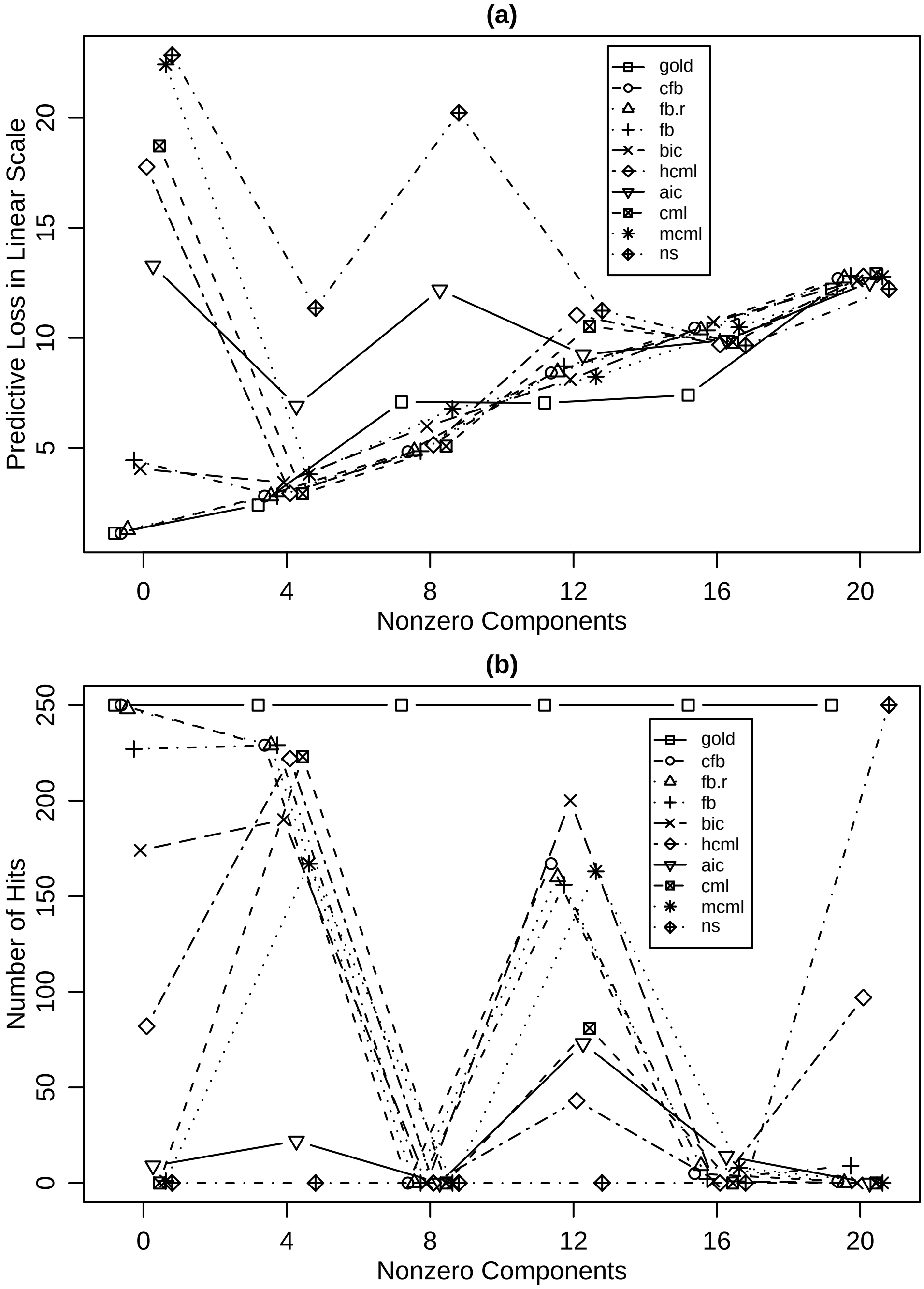

Figure 7.3: Comparison: Poisson, $\rho = 0, n = 200, p = 20$

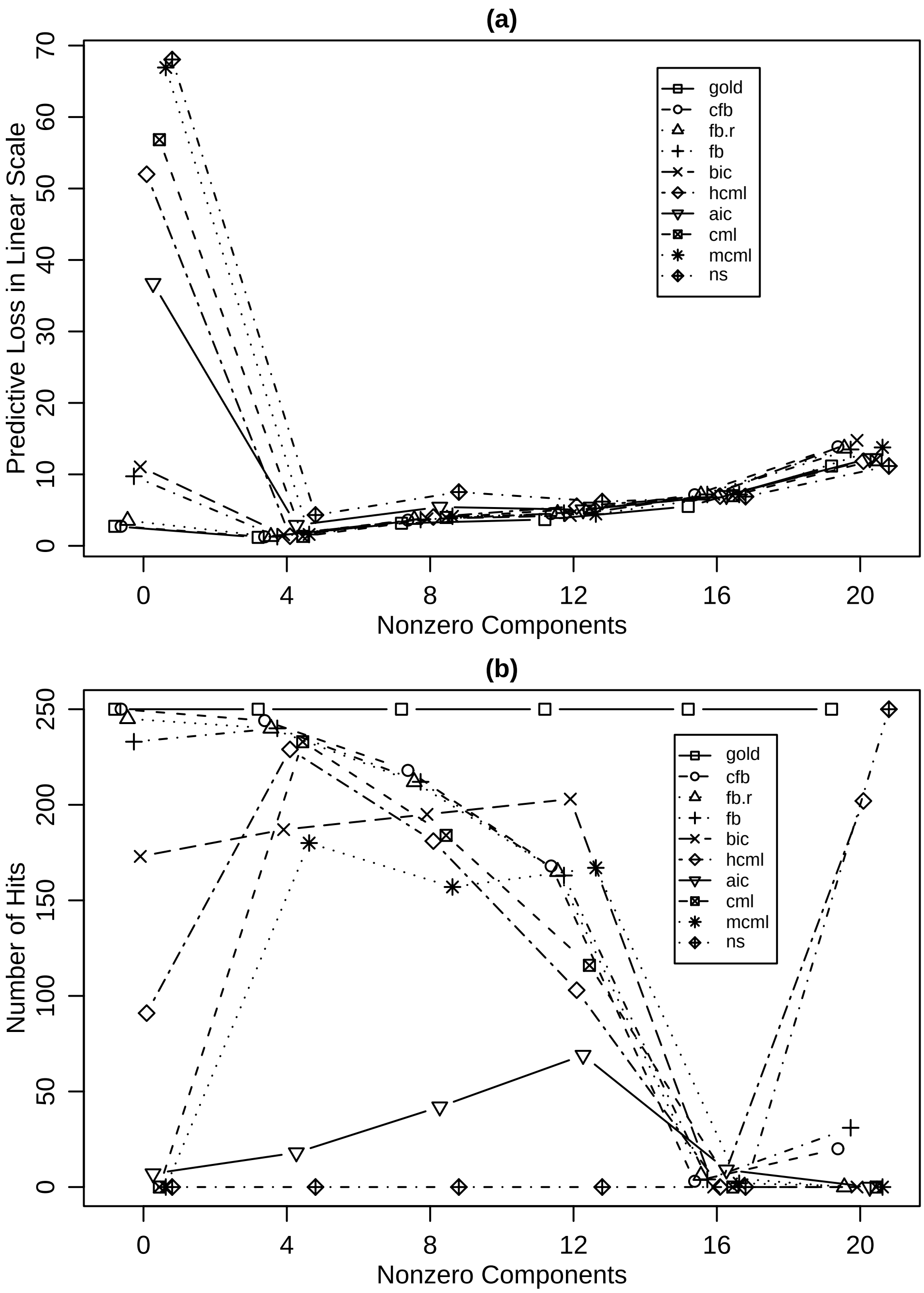

Figure 7.4: Comparison: Poisson, $\rho = 0.5, n = 200, p = 20$

7.4 Logistic Models

I report simulation results based on Logistic models in this section. For each Y_i, under the Logistic model we have :

$$f(y_i|\mu_i) = \mu_i^{y_i}(1-\mu_i)^{1-y_i} \text{ for } y_i = 0, 1$$

where μ_i is the mean of Y_i, also the probability of Y_i being 1. We know the distribution of Y_i is $B(1, \mu_i)$. From Table (1.1), we have $\phi = 1$, $b(\theta_i) = \log(1 + e^{\theta_i})$ and $c(y_i, \phi) = 0$. Also, under the canonical link function, we have $\boldsymbol{\mu} = \frac{1}{1+\exp(-\mathbf{X}\boldsymbol{\beta})}$. I define the predictive loss in the fitted scale since a fitted value is between 0 and 1 and has small variation compared to the one defined in the linear scale.

Table (7.11) to Table (7.16) along with Figure (7.5), Figure (7.6) summarize the results by n and p. For the case $n = 200, p = 20$, Table (7.17) to Table (7.20) along with Figure (7.7), Figure (7.8) describe the results by number of nonzero components (for other n and p, please refer to Appendix C.2). All these tables and figures are parallel to those given in last section for Poisson models.

Based on predictive loss, my findings in overall performance are as follows.

1. Again, the three FB criteria are significantly better than the three EB criteria.

2. Among the FB criteria, $FB.r$ works best. At most times, it is better than CFB, and CFB performs better than or at least comparable to FB. $FB.r$ improves FB a lot. Also, $FB.r$ and CFB are usually better than AIC and BIC. Only when p equals 50, is $FB.r$ worse than BIC, but still better or comparable to AIC while CFB is worse than both. FB sometimes works better, sometimes worse than BIC or AIC. Note, unlike under Poisson models, AIC sometimes performs better than BIC. In addition, it is worth mentioning that there were many warnings of convergence problems when running simulations for $p = 50$ under Logistic models. The GLM estimation algorithm sometimes does not converge when q is large. It appeals that BIC is pretty robust to this problem, while our $FB.r$ and CFB are sensitive. This would not be surprising, since the penalties of adding a variable in $FB.r$ and CFB depend heavily on the parameter estimates while BIC has a data-independent penalty.

3. The EB criteria has very similar performance to the nonselection method. Basically speaking, they are worse than other criteria.

Based on percent of hits, my findings in overall performance are listed below.

1. The three FB criteria are better than other criteria in hitting the right models. Among them, CFB works best; FB and $FB.r$ look similar to each other. However, CFB works very well when the true model is either null or full and FB works very well when the true model is full, which

brings up their percent of hits. When the true model is not full or null, they work poorly and consistently worse than $FB.r$ and BIC.

2. Among the EB criteria, $MCML$ and $HCML$ are very similar to the nonselection method and can only work well when the true model is full, while CML usually chooses wrong models. If the true model is not full, they are worse than BIC and AIC.

3. Generally speaking, only $FB.r$ and BIC can help in choosing right models when the true model is not full, but they do not help a lot for Logistic models, especially for a model with a large q.

By looking at the relative performance of criteria at different q, I find that there is no criterion that works uniformly better than others at any number of nonzero components. $FB.r$ and CFB work well when the true model is null or full. At other q's, they have no consistent pattern. Even when they are worse than some criterion, the difference in predictive loss usually is small. So they often beat others in the overall performance. Other criteria often have very bad performance in predictive loss at some certain q, for example, BIC is much worse than others when q is close to p, AIC and the EB criteria are very bad when $q = 0$, which pulls down their overall performance.

For Logistic models, I conclude that generally, $FB.r$ and CFB have better overall performance than the traditional criteria. This might hold when p is not large. If p is large and n is not large, BIC might have better performance since it seems resistant to the convergence problem. Among the three FB criteria, CFB and $FB.r$ are better than FB in predictive loss. The EB criteria have similar performance to the method of always picking the full model and they are worse than the traditional ones. Here, no criterion can achieve the top performance at any number of nonzero components.

Table 7.11: Average Predictive Loss: Logistic, $\rho = 0$

p	n	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
10	100	1.15	1.91	1.98	2.23	2.08	2.12	2.23	2.27	2.28	2.28
		0.02	*0.03*	*0.03*	*0.04*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
10	200	1.14	1.81	1.99	1.99	1.87	1.99	2.13	2.12	2.16	2.16
		0.02	*0.03*	*0.03*	*0.04*	*0.03*	*0.03*	*0.02*	*0.02*	*0.02*	*0.02*
10	500	1.12	1.47	1.47	2.07	1.85	1.74	2.08	2.09	2.13	2.13
		0.02	*0.03*	*0.03*	*0.04*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
20	100	2.27	3.46	3.51	3.93	4.03	4.51	4.58	4.55	4.60	4.60
		0.05	*0.06*	*0.06*	*0.06*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*
20	200	2.14	3.59	3.73	4.95	4.15	4.26	4.41	4.47	4.49	4.49
		0.04	*0.05*	*0.05*	*0.08*	*0.04*	*0.05*	*0.04*	*0.04*	*0.04*	*0.04*
20	500	2.07	2.65	2.82	3.50	3.57	3.48	4.10	3.97	4.25	4.25
		0.04	*0.05*	*0.05*	*0.08*	*0.04*	*0.05*	*0.04*	*0.04*	*0.04*	*0.04*
50	200	5.33	9.52	10.96	8.36	9.50	11.47	11.45	11.30	11.46	11.47
		0.07	*0.09*	*0.07*	*0.08*	*0.06*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.12: Percent of Hits: Logistic, $\rho = 0$

p	n	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
10	100	1.00	0.24	0.33	0.19	0.11	0.26	0.00	0.17	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
10	200	1.00	0.31	0.36	0.26	0.12	0.28	0.00	0.08	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
10	500	1.00	0.35	0.38	0.26	0.22	0.33	0.00	0.18	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	100	1.00	0.18	0.32	0.09	0.01	0.20	0.00	0.12	0.13	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	200	1.00	0.20	0.33	0.16	0.03	0.23	0.00	0.17	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	500	1.00	0.35	0.44	0.27	0.03	0.34	0.00	0.04	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
50	200	1.00	0.10	0.14	0.03	0.00	0.09	0.01	0.00	0.04	0.09
		0.00	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.13: Average Predictive Loss: Logistic, $\rho = 0.5$

p	n	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
10	100	1.19	1.64	1.57	1.88	2.03	2.08	2.29	2.33	2.36	2.36
		0.02	*0.03*	*0.03*	*0.04*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
10	200	1.17	1.65	1.66	2.04	2.01	2.03	2.25	2.26	2.30	2.30
		0.02	*0.03*	*0.03*	*0.04*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
10	500	1.12	1.83	1.96	2.15	1.98	1.96	2.17	2.18	2.21	2.21
		0.02	*0.04*	*0.04*	*0.04*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*
20	100	2.29	3.83	3.77	3.99	4.11	4.62	4.67	4.64	4.68	4.69
		0.04	*0.06*	*0.05*	*0.06*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*
20	200	2.12	2.98	3.22	3.63	3.78	4.24	4.39	4.42	4.46	4.46
		0.04	*0.05*	*0.05*	*0.06*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*	*0.04*
20	500	2.11	3.24	3.61	3.56	3.50	3.83	4.14	4.10	4.30	4.30
		0.04	*0.05*	*0.05*	*0.06*	*0.04*	*0.05*	*0.04*	*0.04*	*0.04*	*0.04*
50	200	5.46	8.52	10.82	6.35	9.03	11.65	11.63	11.45	11.64	11.65
		0.07	*0.10*	*0.08*	*0.06*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*	*0.05*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.14: Percent of Hits: Logistic, $\rho = 0.5$

p	n	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
10	100	1.00	0.19	0.32	0.15	0.06	0.25	0.00	0.16	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
10	200	1.00	0.20	0.33	0.15	0.06	0.29	0.00	0.04	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
10	500	1.00	0.40	0.46	0.39	0.18	0.38	0.01	0.13	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	100	1.00	0.18	0.32	0.10	0.01	0.19	0.00	0.15	0.15	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	200	1.00	0.17	0.33	0.12	0.02	0.22	0.00	0.17	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.01*
20	500	1.00	0.25	0.35	0.22	0.06	0.28	0.00	0.02	0.17	0.17
		0.00	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.01*
50	200	1.00	0.09	0.13	0.04	0.00	0.09	0.01	0.00	0.05	0.09
		0.00	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.15: Paired Comparison in Predictive Loss: Logistic, $\rho = 0$

Type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
fb.r-cfb	-0.07***	-0.17***	0.00	-0.05	-0.14***	-0.17***	-1.45***
fb.r-fb	-0.21***	-0.18***	-0.27***	-1.05***	-0.67***	-0.83***	-1.95***
cfb-fb	-0.14***	-0.00	-0.27***	-1.00***	-0.53***	-0.66***	-0.50***
fb.r-bic	-0.32***	-0.18***	-0.59***	-0.47***	-1.35***	-0.85***	1.16***
cfb-bic	-0.25***	-0.01	-0.60***	-0.42***	-1.22***	-0.68***	2.61***
fb-bic	-0.11**	-0.00	-0.33***	0.58***	-0.69***	-0.02	3.11***
fb.r-aic	-0.17***	-0.06*	-0.37***	-0.57***	-0.56***	-0.92***	0.01
cfb-aic	-0.10***	0.11***	-0.38***	-0.52***	-0.42***	-0.75***	1.46***
fb-aic	0.04*	0.12***	-0.11***	0.48***	0.11**	-0.09**	1.96***

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.16: Paired Comparison in Predictive Loss: Logistic, $\rho = 0.5$

Type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
fb.r-cfb	0.07***	-0.02	-0.13***	0.07	-0.24***	-0.37***	-2.29***
fb.r-fb	-0.44***	-0.39***	-0.13***	-0.78***	-1.26***	-0.59***	-3.13***
cfb-fb	-0.51***	-0.37***	0.00	-0.85***	-1.02***	-0.22***	-0.84***
fb.r-bic	-0.24***	-0.39***	-0.32***	-0.15***	-0.65***	-0.32***	2.18***
cfb-bic	-0.31***	-0.37***	-0.18***	-0.22***	-0.41***	0.05	4.47***
fb-bic	0.20***	-0.00	-0.19***	0.63***	0.61***	0.27***	5.31***
fb.r-aic	-0.39***	-0.37***	-0.15***	-0.27***	-0.80***	-0.27***	-0.51***
cfb-aic	-0.46***	-0.35***	-0.01	-0.34***	-0.56***	0.11*	1.78***
fb-aic	0.05*	0.02	-0.02	0.51***	0.46***	0.33***	2.62***

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table 7.17: Average Predictive Loss by Nonzero Components
Logistic, $\rho = 0, n = 200, p = 20$

q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
0	0.25	0.29	0.33	1.09	3.26	3.96	4.82	5.22	5.32	5.32
	0.02	*0.04*	*0.06*	*0.08*	*0.11*	*0.20*	*0.14*	*0.11*	*0.10*	*0.10*
4	1.08	4.72	5.44	3.66	3.90	4.99	4.98	4.98	4.99	4.99
	0.05	*0.09*	*0.07*	*0.11*	*0.11*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*
8	1.65	3.83	4.18	3.49	3.42	4.18	4.18	4.18	4.18	4.18
	0.05	*0.10*	*0.09*	*0.12*	*0.10*	*0.09*	*0.08*	*0.09*	*0.09*	*0.09*
12	2.49	4.25	4.13	5.38	4.26	4.13	4.13	4.13	4.13	4.13
	0.06	*0.08*	*0.08*	*0.09*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
16	3.38	4.39	4.29	8.62	5.12	4.29	4.31	4.29	4.29	4.29
	0.07	*0.08*	*0.08*	*0.14*	*0.10*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
20	4.00	4.09	4.00	7.44	4.93	4.00	4.02	4.00	4.00	4.00
	0.07	*0.07*	*0.07*	*0.11*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*

Table 7.18: No. of Hits by Nonzero Components
Logistic, $\rho = 0, n = 200, p = 20$

q	fb.r	cfb	bic	aic	fb	cml	mcml	hcml
0	248	248	149	5	95	0	0	0
4	10	0	22	5	0	0	0	0
8	23	0	64	29	0	0	0	0
12	0	0	1	0	0	0	0	0
16	1	0	0	2	0	0	0	0
20	15	250	0	0	250	0	249	250

Table 7.19: Average Predictive Loss by Nonzero Components
Logistic, $\rho = 0.5, n = 200, p = 20$

q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
0	0.24	0.35	0.33	1.04	3.31	4.16	5.00	5.25	5.37	5.37
	0.02	*0.05*	*0.06*	*0.08*	*0.11*	*0.19*	*0.13*	*0.11*	*0.10*	*0.10*
2	1.11	2.32	2.58	2.65	3.60	4.78	4.85	4.88	4.90	4.90
	0.04	*0.05*	*0.10*	*0.06*	*0.09*	*0.10*	*0.09*	*0.09*	*0.09*	*0.09*
4	1.55	3.87	4.04	4.05	3.53	4.04	4.04	4.04	4.04	4.04
	0.04	*0.09*	*0.07*	*0.11*	*0.09*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
6	2.36	2.29	3.99	1.58	2.86	4.08	4.08	4.01	4.08	4.08
	0.06	*0.11*	*0.09*	*0.08*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
8	3.36	4.76	4.28	5.86	4.68	4.28	4.28	4.28	4.28	4.28
	0.08	*0.09*	*0.08*	*0.09*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
10	4.07	4.26	4.07	6.59	4.67	4.07	4.08	4.08	4.07	4.07
	0.08	*0.09*	*0.08*	*0.10*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*

Table 7.20: No. of Hits by Nonzero Components
Logistic, $\rho = 0.5, n = 200, p = 20$

q	fb.r	cfb	bic	aic	fb	cml	mcml	hcml
0	245	248	160	8	86	0	0	0
4	0	0	0	0	0	0	0	0
8	11	0	16	21	0	0	0	0
12	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0
20	2	250	0	0	250	0	249	250

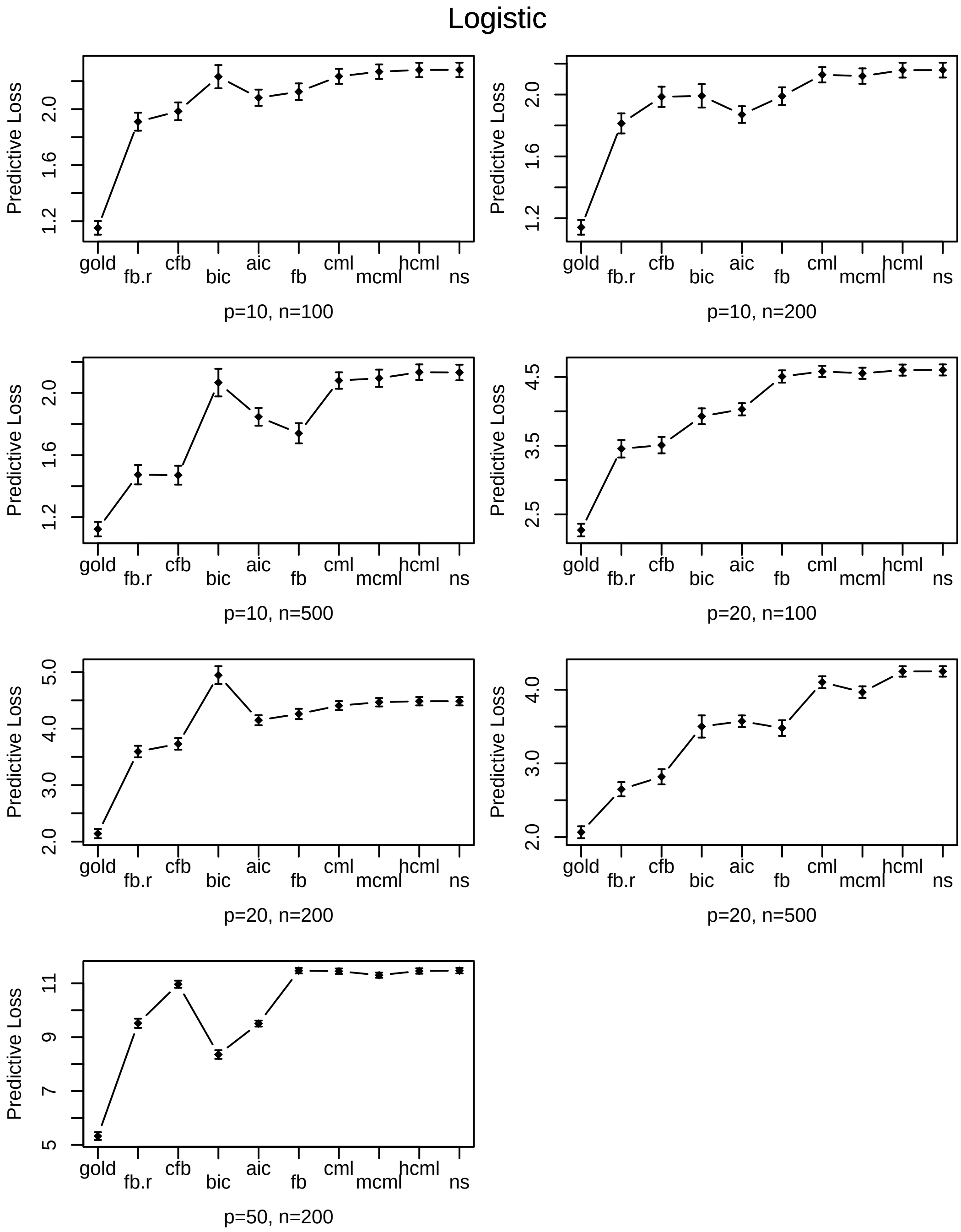

Figure 7.5: Average Predictive Loss With 95% CI: Logistic, $\rho = 0$

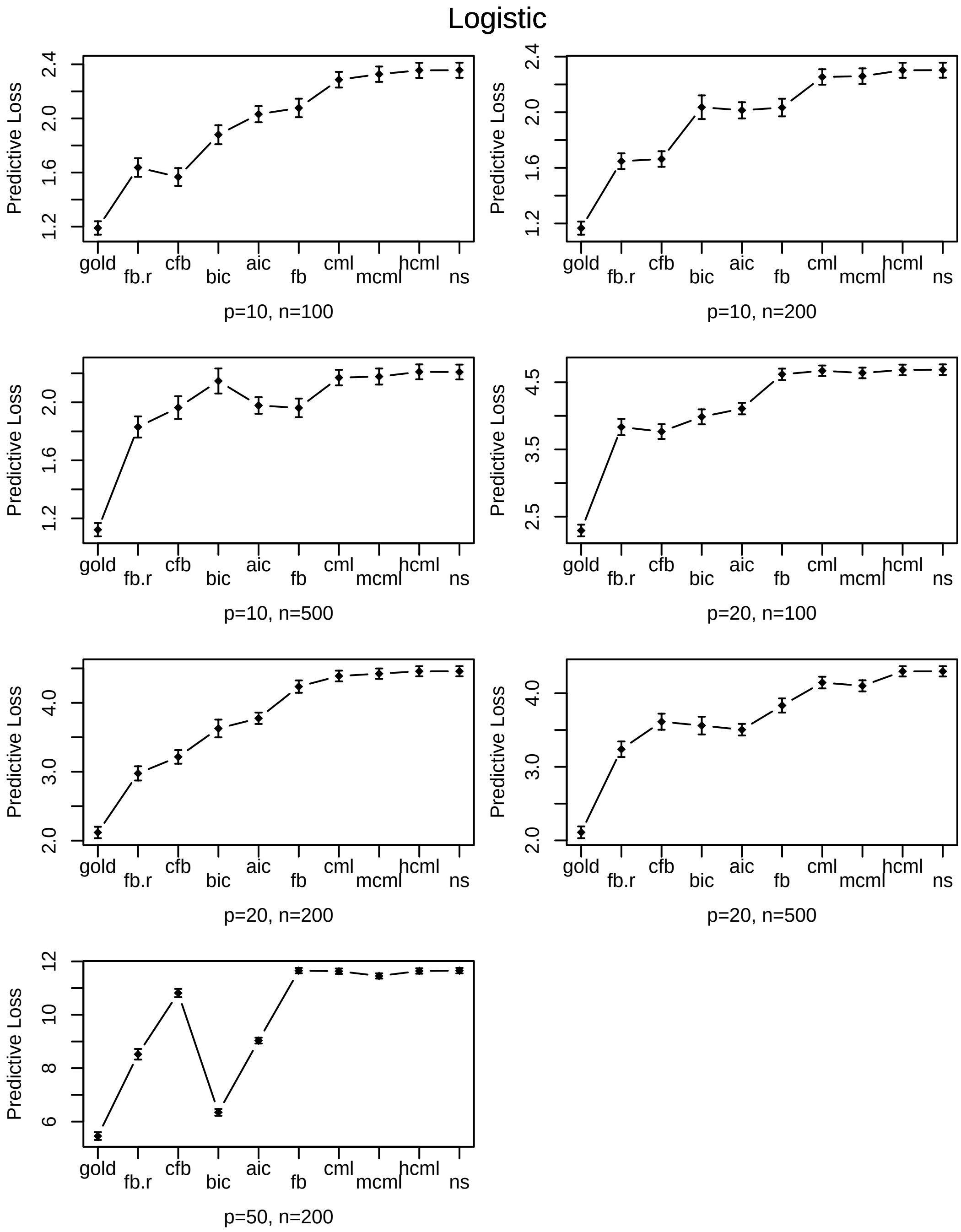

Figure 7.6: Average Predictive Loss With 95% CI: Logistic, $\rho = 0.5$

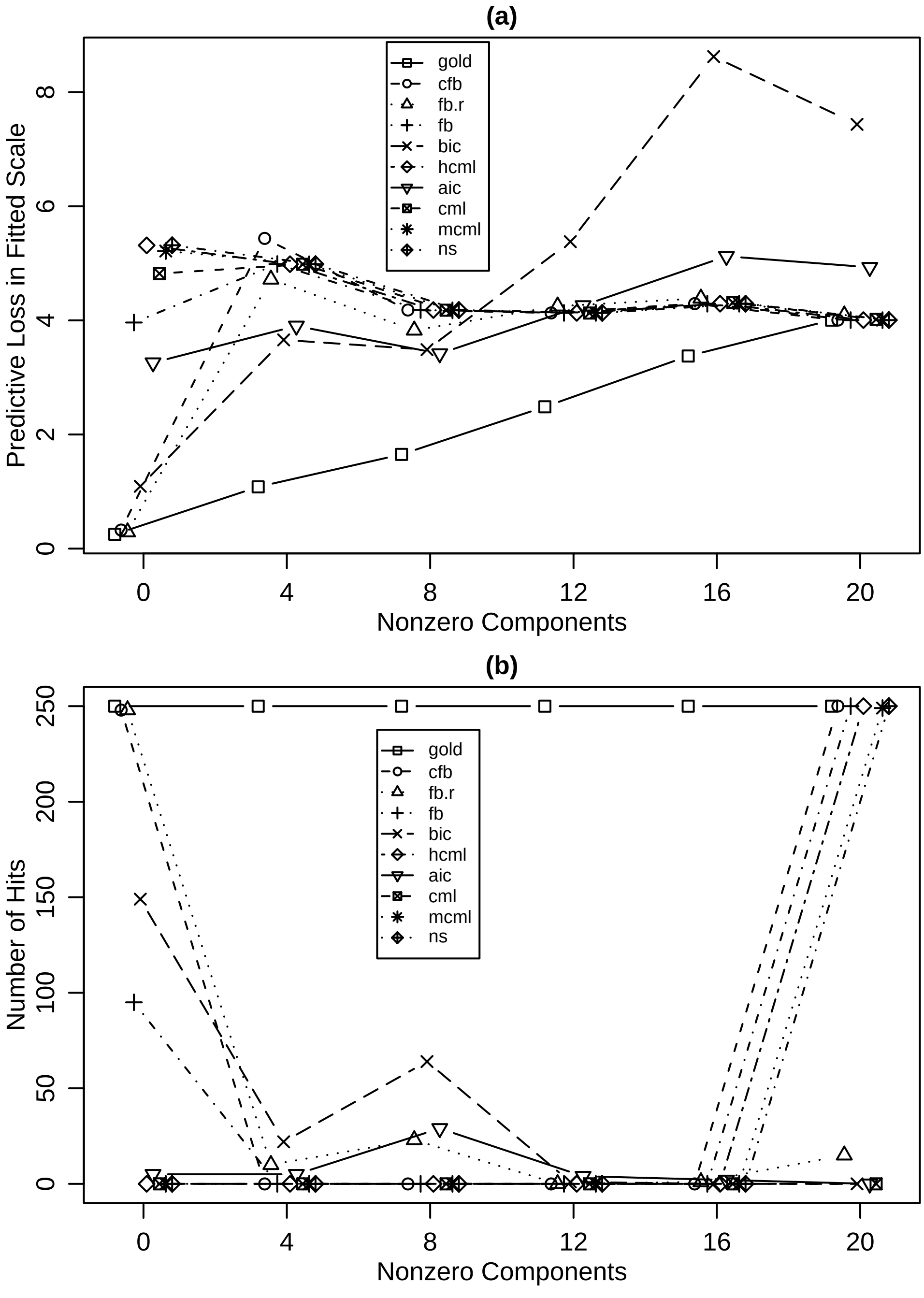

Figure 7.7: Comparison: Logistic, $\rho = 0, n = 200, p = 20$

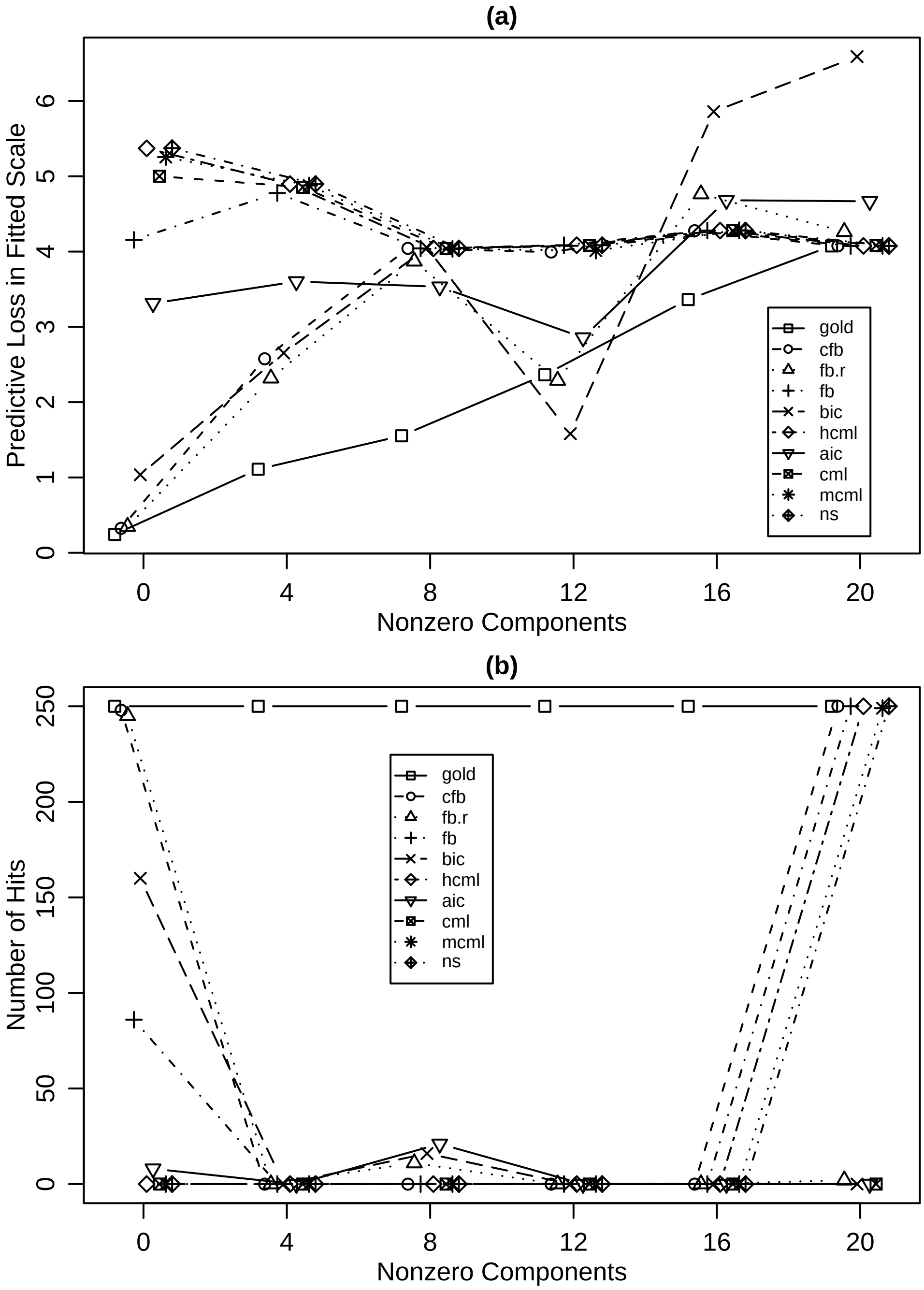

Figure 7.8: Comparison: Logistic, $\rho = 0.5, n = 200, p = 20$

7.5 Normal Linear Models

I report simulation results based on normal linear models in this section.

Table (7.21) to Table (7.26) along with Figure (7.9), Figure (7.10) summarize the results by n and p. For the case $n = 200, p = 20$, Table (7.27) to Table (7.30) along with Figure (7.11), Figure (7.12) describe the results by different q (for other n and p, please refer to Appendix C.3).

Based on predictive loss, my findings in overall performance are as follows.

1. The three FB criteria are significantly better than the three EB criteria.
2. All the FB criteria are uniformly better than AIC. CFB and $FB.r$ are uniformly better than BIC, and at most times FB is better than BIC. Among the three FB criteria, CFB works better than the other two; $FB.r$ is often better than FB.
3. All the three EB criteria are worse than BIC. Often they are slightly worse than or similar to AIC. Among themselves, they are all at the same level and none of them is consistently better or worse than others.

Based on percent of hits, my findings in overall performance are listed below.

1. The FB criteria work much better than the EB criteria.
2. The FB criteria perform significantly better than AIC, and are better than BIC. Among themselves, CFB is the best. Unlike in predictive

loss, $FB.r$ is worse than or comparable to FB in hitting the right model. The reason is the same as discussed for Poisson models, that is, FB tends to choose the full model so it works better than $FB.r$ when the true model is full, which brings up its percent of hits.

3. Among the EB criteria, $HCML$ works best and much outperforms AIC. Sometimes it is even better than BIC. The other two, $MCML$ and CML, are worse than BIC, but at most times are better than or comparable to AIC. Obviously, $HCML$ improves CML notably and consistently in hitting the right model while $MCML$ does not.

By looking at the relative performance of criteria at different q, I find that CFB usually achieves the top performance at any number of nonzero components. Also, $FB.r$ is similar to CFB in performance except that it is sometimes a bit worse when the true model is full.

For normal linear models, I conclude that the FB criteria are consistently better than the EB or traditional criteria. Among the three FB criteria, CFB improves FB in both predictive loss and percent of hits. $FB.r$ often improves FB in predictive loss. The EB criteria usually do not exceed BIC in variable selection. In addition, CFB and $FB.r$ usually achieve the top performance at any number of nonzero components. Not surprisingly, these conclusions are similar to those under Poisson models.

Table 7.21: Average Predictive Loss: Normal, $\rho = 0$

p	n	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
10	100	6.21	8.50	8.56	9.02	9.34	9.81	10.28	10.30	10.51	11.24
		0.13	*0.17*	*0.17*	*0.16*	*0.19*	*0.14*	*0.14*	*0.14*	*0.14*	*0.12*
10	200	6.20	7.57	7.44	8.28	7.73	9.08	9.73	9.38	9.87	11.28
		0.13	*0.17*	*0.16*	*0.17*	*0.18*	*0.14*	*0.15*	*0.15*	*0.16*	*0.13*
10	500	6.06	7.19	7.34	7.64	7.77	9.24	9.02	9.17	8.97	11.07
		0.13	*0.15*	*0.15*	*0.16*	*0.17*	*0.14*	*0.15*	*0.15*	*0.15*	*0.12*
20	100	11.01	15.76	16.24	16.76	20.11	18.70	20.41	19.79	20.22	21.07
		0.21	*0.27*	*0.30*	*0.27*	*0.36*	*0.22*	*0.20*	*0.19*	*0.21*	*0.17*
20	200	11.04	14.79	14.68	15.95	23.72	19.36	19.56	19.97	19.79	21.25
		0.21	*0.27*	*0.28*	*0.27*	*0.51*	*0.24*	*0.23*	*0.23*	*0.22*	*0.17*
20	500	11.02	14.03	14.21	14.70	18.49	18.09	17.61	18.77	17.69	21.05
		0.21	*0.26*	*0.27*	*0.26*	*0.37*	*0.21*	*0.23*	*0.23*	*0.24*	*0.17*
50	200	25.99	35.88	39.95	38.61	43.80	43.77	49.85	46.85	50.09	50.87
		0.33	*0.39*	*0.50*	*0.39*	*0.54*	*0.29*	*0.24*	*0.24*	*0.23*	*0.19*

Table 7.22: Percent of Hits: Normal, $\rho = 0$

p	n	cfb	fb.r	fb	bic	aic	cml	mcml	hcml
10	100	0.46	0.38	0.46	0.31	0.15	0.12	0.15	0.29
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
10	200	0.56	0.44	0.55	0.42	0.13	0.21	0.18	0.40
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
10	500	0.69	0.61	0.68	0.60	0.25	0.31	0.38	0.51
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
20	100	0.45	0.29	0.42	0.19	0.02	0.01	0.01	0.23
		0.01	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*
20	200	0.53	0.48	0.48	0.34	0.08	0.09	0.11	0.28
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
20	500	0.60	0.48	0.57	0.39	0.08	0.20	0.16	0.39
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
50	200	0.22	0.13	0.21	0.07	0.00	0.00	0.00	0.11
		0.01	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*

Table 7.23: Average Predictive Loss: Normal, $\rho = 0.5$

p	n	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
10	100	5.93	7.88	8.34	8.12	9.53	9.57	9.54	9.96	10.01	11.06
		0.13	*0.19*	*0.20*	*0.18*	*0.21*	*0.16*	*0.16*	*0.16*	*0.15*	*0.13*
10	200	5.92	7.11	7.22	7.81	7.61	9.11	8.96	9.27	9.02	10.97
		0.13	*0.15*	*0.15*	*0.16*	*0.16*	*0.14*	*0.15*	*0.14*	*0.16*	*0.13*
10	500	6.04	7.35	7.48	7.83	9.10	9.42	9.36	9.61	9.06	11.10
		0.13	*0.16*	*0.16*	*0.16*	*0.21*	*0.14*	*0.15*	*0.16*	*0.15*	*0.12*
20	100	11.00	13.90	14.20	15.19	15.76	17.06	20.09	18.51	20.12	20.89
		0.21	*0.27*	*0.29*	*0.27*	*0.30*	*0.20*	*0.20*	*0.19*	*0.20*	*0.17*
20	200	11.04	12.52	13.38	13.36	15.85	17.89	18.65	18.77	19.77	21.45
		0.21	*0.27*	*0.29*	*0.26*	*0.34*	*0.21*	*0.24*	*0.21*	*0.23*	*0.17*
20	500	11.01	15.20	15.52	14.97	20.29	18.04	18.51	19.05	17.69	20.85
		0.21	*0.32*	*0.33*	*0.30*	*0.45*	*0.23*	*0.26*	*0.25*	*0.24*	*0.17*
50	200	25.79	36.04	39.10	37.83	41.19	42.78	49.52	45.01	49.95	50.94
		0.33	*0.40*	*0.49*	*0.40*	*0.50*	*0.28*	*0.25*	*0.25*	*0.24*	*0.19*

Table 7.24: Percent of Hits: Normal, $\rho = 0.5$

p	n	cfb	fb.r	fb	bic	aic	cml	mcml	hcml
10	100	0.51	0.38	0.47	0.34	0.14	0.11	0.14	0.28
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
10	200	0.47	0.33	0.44	0.28	0.09	0.14	0.11	0.37
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
10	500	0.76	0.73	0.70	0.68	0.38	0.32	0.51	0.50
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
20	100	0.30	0.29	0.29	0.20	0.03	0.02	0.02	0.22
		0.01	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*
20	200	0.42	0.29	0.42	0.23	0.02	0.07	0.03	0.25
		0.01	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*
20	500	0.53	0.51	0.52	0.33	0.14	0.18	0.19	0.39
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*
50	200	0.31	0.21	0.28	0.12	0.00	0.01	0.00	0.12
		0.01	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.01*

Table 7.25: Paired Comparison in Predictive Loss: Normal, $\rho = 0$

Type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
cfb-fb.r	-0.06	0.13**	-0.14***	-0.48***	0.11	-0.18**	-4.07***
cfb-fb	-0.52***	-0.71***	-0.45***	-1.00***	-1.16***	-0.67***	-2.73***
fb.r-fb	-0.46***	-0.84***	-0.30***	-0.52**	-1.27***	-0.49***	1.34***
cfb-bic	-0.84***	-0.17	-0.58***	-4.35***	-8.93***	-4.46***	-7.92***
fb.r-bic	-0.77***	-0.30**	-0.44***	-3.88***	-9.04***	-4.28***	-3.85***
fb-bic	-0.32*	0.54***	-0.13	-3.36***	-7.77***	-3.79***	-5.19***
cfb-aic	-1.31***	-1.51***	-2.05***	-2.94***	-4.57***	-4.05***	-7.89***
fb.r-aic	-1.25***	-1.64***	-1.91***	-2.46***	-4.68***	-3.88***	-3.82***
fb-aic	-0.79***	-0.80***	-1.60***	-1.94***	-3.41***	-3.38***	-5.16***

Table 7.26: Paired Comparison in Predictive Loss: Normal, $\rho = 0.5$

Type	p=10 n=100	p=10 n=200	p=10 n=500	p=20 n=100	p=20 n=200	p=20 n=500	p=50 n=200
cfb-fb.r	-0.45***	-0.12**	-0.14***	-0.30*	-0.86***	-0.33***	-3.05***
cfb-fb	-0.24*	-0.70***	-0.48***	-1.29***	-0.83***	0.22	-1.79***
fb.r-fb	0.22	-0.59***	-0.34***	-1.00***	0.02	0.55***	1.26***
cfb-bic	-1.64***	-0.50***	-1.75***	-1.86***	-3.33***	-5.10***	-5.15***
fb.r-bic	-1.19***	-0.39***	-1.61***	-1.56***	-2.47***	-4.77***	-2.10***
fb-bic	-1.40***	0.20*	-1.27***	-0.57*	-2.49***	-5.32***	-3.36***
cfb-aic	-1.68***	-2.01***	-2.08***	-3.16***	-5.36***	-2.85***	-6.74***
fb.r-aic	-1.23***	-1.89***	-1.94***	-2.86***	-4.50***	-2.52***	-3.69***
fb-aic	-1.45***	-1.30***	-1.60***	-1.86***	-4.53***	-3.07***	-4.95***

Table 7.27: Average Predictive Loss by Nonzero Components
normal, $\rho = 0, n = 200, p = 20$

q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
0	0.94	0.94	1.03	5.27	4.07	12.58	17.66	20.88	16.94	21.23
	0.08	*0.08*	*0.12*	*0.70*	*0.31*	*0.44*	*0.67*	*0.44*	*0.74*	*0.42*
4	5.03	6.94	7.05	7.62	7.66	14.56	12.99	13.01	17.01	21.45
	0.21	*0.35*	*0.34*	*0.43*	*0.34*	*0.46*	*0.70*	*0.48*	*0.68*	*0.43*
8	9.01	17.58	14.92	19.58	13.37	16.60	21.52	16.31	21.54	21.54
	0.26	*0.61*	*0.55*	*0.54*	*0.54*	*0.43*	*0.41*	*0.44*	*0.42*	*0.42*
12	13.52	21.36	20.92	21.31	26.71	19.93	21.29	19.92	21.31	21.31
	0.33	*0.42*	*0.53*	*0.42*	*0.76*	*0.51*	*0.42*	*0.50*	*0.42*	*0.42*
16	16.98	21.16	22.65	21.16	39.61	25.12	21.25	24.05	21.16	21.16
	0.38	*0.42*	*0.52*	*0.42*	*0.71*	*0.55*	*0.43*	*0.53*	*0.42*	*0.42*
20	20.79	20.79	21.52	20.79	50.93	27.39	22.64	25.61	20.79	20.79
	0.42	*0.42*	*0.45*	*0.42*	*0.86*	*0.56*	*0.40*	*0.55*	*0.42*	*0.42*

Table 7.28: No. of Hits by Nonzero Components
normal, $\rho = 0, n = 200, p = 20$

q	cfb	fb.r	fb	bic	aic	cml	mcml	hcml
0	250	249	216	157	11	0	1	74
4	200	194	191	171	18	141	43	90
8	99	110	62	154	27	0	44	0
12	0	36	0	29	43	0	41	0
16	0	11	0	1	11	0	15	0
20	250	121	250	0	16	0	27	250

Table 7.29: Average Predictive Loss by Nonzero Components
normal, $\rho = 0.5, n = 200, p = 20$

q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
0	0.95 *0.08*	0.95 *0.08*	1.00 *0.09*	5.54 *0.72*	3.43 *0.29*	12.30 *0.46*	17.41 *0.69*	20.74 *0.46*	16.65 *0.76*	21.16 *0.43*
4	5.13 *0.21*	7.54 *0.46*	7.73 *0.41*	8.30 *0.52*	8.20 *0.41*	15.40 *0.48*	16.80 *0.73*	14.35 *0.51*	20.17 *0.62*	22.62 *0.45*
8	9.48 *0.27*	11.68 *0.37*	11.91 *0.36*	12.45 *0.43*	12.14 *0.34*	17.38 *0.39*	18.30 *0.58*	16.19 *0.41*	20.50 *0.52*	22.25 *0.41*
12	12.85 *0.33*	9.53 *0.40*	9.58 *0.37*	10.93 *0.49*	9.95 *0.34*	16.09 *0.42*	17.80 *0.59*	15.38 *0.43*	19.71 *0.51*	21.09 *0.42*
16	17.15 *0.35*	20.93 *0.39*	20.68 *0.43*	20.93 *0.39*	27.18 *0.72*	20.26 *0.45*	20.90 *0.39*	20.11 *0.44*	20.93 *0.39*	20.93 *0.39*
20	20.67 *0.38*	24.52 *0.58*	29.39 *0.51*	22.01 *0.49*	34.22 *0.52*	25.88 *0.39*	20.69 *0.38*	25.82 *0.39*	20.67 *0.38*	20.67 *0.38*

Table 7.30: No. of Hits by Nonzero Components
normal, $\rho = 0.5, n = 200, p = 20$

q	cfb	fb.r	fb	bic	aic	cml	mcml	hcml
0	250	249	213	178	10	0	0	79
4	201	188	186	167	18	103	41	53
8	0	0	0	0	1	0	1	0
12	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0
20	183	1	226	0	0	0	0	250

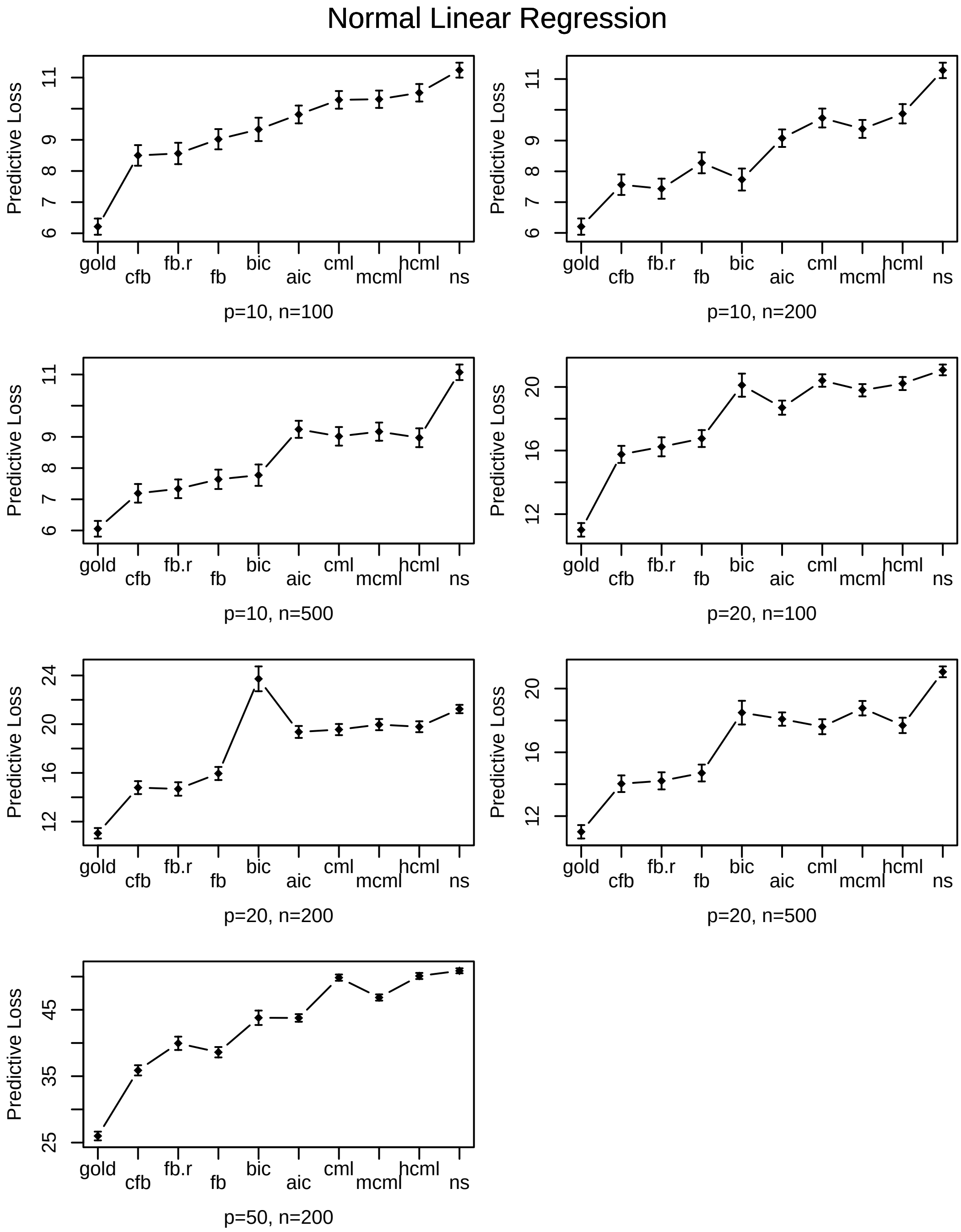

Figure 7.9: Average Predictive Loss With 95% CI: Normal, $\rho = 0$

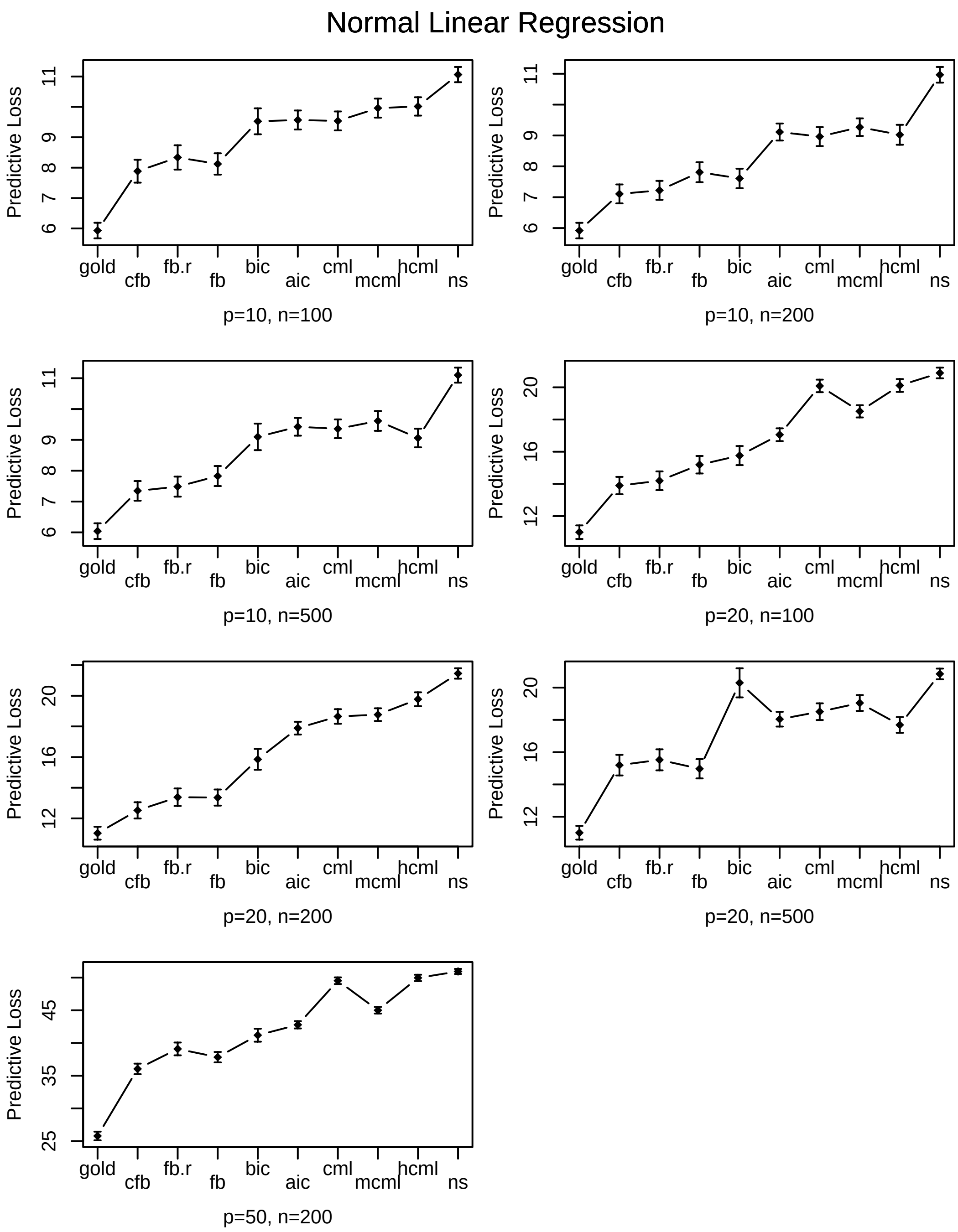

Figure 7.10: Average Predictive Loss With 95% CI: Normal, $\rho = 0.5$

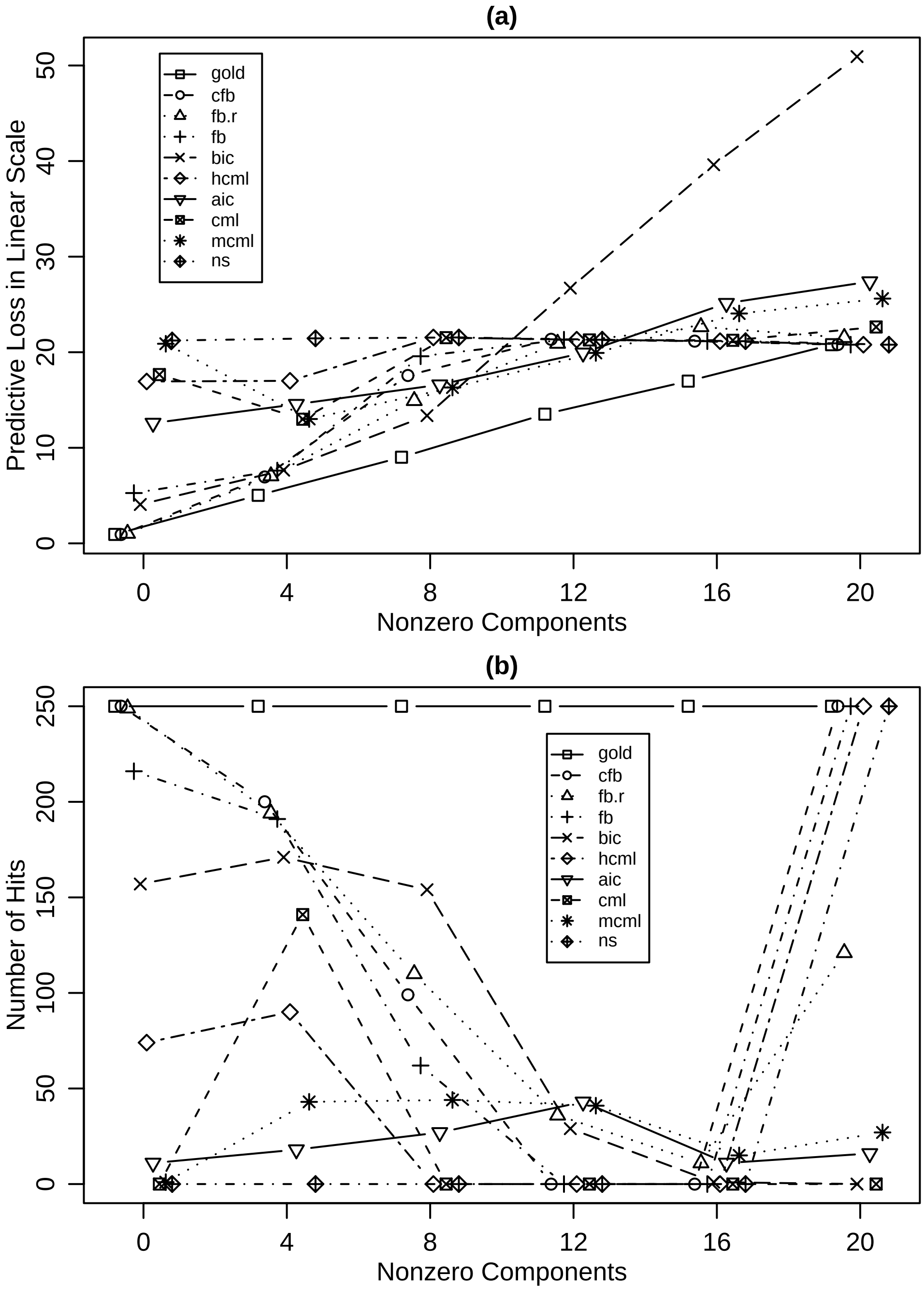

Figure 7.11: Comparison: Normal, $\rho = 0, n = 200, p = 20$

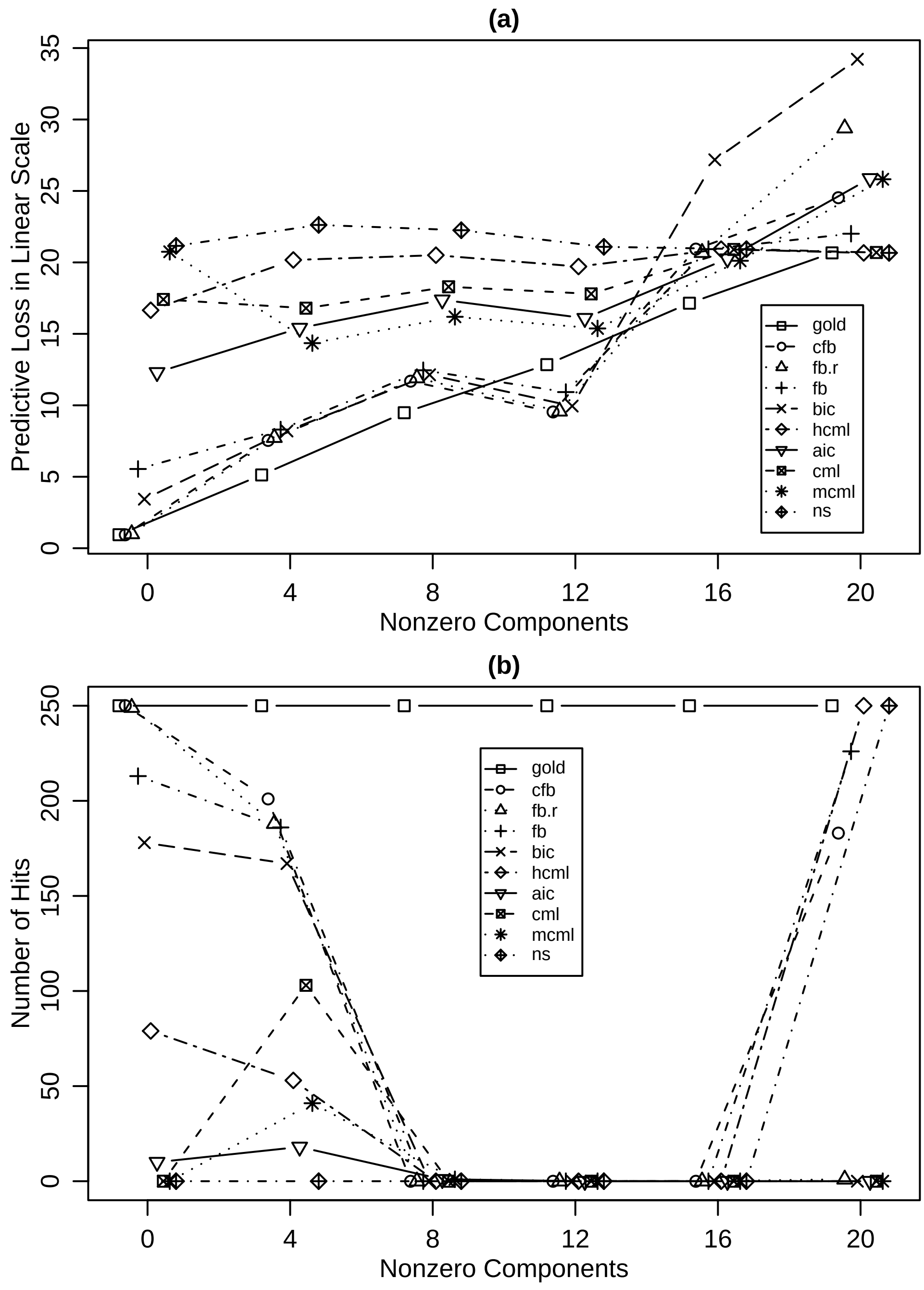

Figure 7.12: Comparison: Normal, $\rho = 0.5, n = 200, p = 20$

7.6 Performance Under The Forward and Backward Procedures

In practice, people usually run the forward, backward or stepwise procedure to select the "best" model based on a criterion using SAS or Splus/R. Instead of choosing a subset first, a criterion is directly applied during a selection procedure and the procedure stops at a final model when the criterion is not increasing /decreasing if adding /dropping a variable. For example, the following steps describe how the forward procedure proceeds using AIC :

1. first fit a null model with no variable in. Then fit a simple GLM regression model for each p potential X variables. Compare the smallest AIC value of all the one-variable models to the AIC value of the null model, if it goes down, then add the corresponding variable $X_{(1)}$ with the smallest AIC value into the model. Otherwise, stop at the null model.

2. Fit all regression models with two X variables, where $X_{(1)}$ is one of the pair. Then compare the smallest AIC value of all the two-variable models to the AIC value of the one-variable model obtained from last step, if it goes down, add the corresponding variable $X_{(2)}$ with the smallest AIC value into the model; otherwise, stop with the one-variable model.

3. Repeat the above procedure, until it stops.

I evaluate the performance of criteria in variable selection under the forward and backward procedures in this section. People might think the performance under a selection procedure should be similar to that obtained in last several sections. However, this is not necessarily true. Let's consider an extreme case. If a criterion always choose the null model when it faces choices that consist of the null model and all the models with only one variable, then under a forward procedure, we would have no chance to access models with more than one variables since the procedure stops at the null model. This cannot happen if we first choose a subset from the whole model space, then apply a criterion to all the members of the subset.

Very often for a single criterion, the number of models visited by the forward/backward procedure is much greater than p. For example, if a forward procedure for a criterion stops at a final model with three variables, the number of models visited for this criterion is $4p-5$. Compared to the method of using a subset of $p+1$ models, this one is time-consuming. So I only run simulations using the forward and backward procedures for Poisson models with $\rho = 0$.

Table (7.31) and (7.32) along with Figure (7.13), summarize the results under the forward procedure (FP) by n and p for Poisson models with $\rho = 0$. My findings in overall performance under FP based on predictive loss and percent of hits are as follows.

1. The three FB criteria work similarly. They are uniformly better than BIC while BIC are much better than AIC.

2. Among the three EB criteria, $HCML$ works best. It is comparable to the FB criteria and it is better than BIC. The other two, CML and

$MCML$ work worse than BIC. Both CML and $MCML$ work uniformly better than AIC in hitting right models. CML does better than AIC in predictive loss when p is 20 or 50. $MCML$ is worse than or at most comparable to AIC in predictive loss.

I carefully check the relative performance of criteria at different q under FP, based on results from different n and p presented in Appendix C.3. Again, I use the case $n = 200, p = 20$ as an example for readers to grasp a basic idea. Table (7.33) and (7.34) along with Figure (7.14) summarize the results under the FP by number of nonzero components for $n = 200, p = 20$. I describe my impression as follows.

1. When the true model is null

   The performance of the criteria is well separated when the true model is null. Basically, the three FB criteria have the top performance and often identify the right model accurately. $HCML$ works slightly worse than the FB criteria, but is still comparable to them. BIC works worse than the first four but most times it can identify the right model. AIC, $MCML$ and CML work poorly: AIC picks up wrong models most of the time; $MCML$ and CML can not identify the right model at all.

2. When the number of nonzero components is less than $p/2$

   The FB criteria usually outperform other criteria in both number of hits and predictive loss. They are slightly better than the EB criteria most of the time, and at the same level as the EB criteria in a few occasions. BIC is worse than the FB criteria, and worse than or comparable to the EB criteria. AIC does not work well and often underperforms the others. Compared to the non-selection method, all the criteria have significantly lower predictive losses.

3. When the number of nonzero components is large than $p/2$

   Generally speaking, all the criteria achieve comparable performance in predictive loss. They do not do much better than the non-selection method. This is reasonable since the full model is not far from the true model when the number of nonzero components is larger than $p/2$. BIC has worse performance here when n is large than when n is small. This is because the penalty of adding a variable for BIC is $\log n$ hence BIC tends to wrongly choose simple models when n is large.

   It is noticed that when the number of nonzero components is large or close to p, sometimes no criterion is able to identify the true model accurately, but their predictive losses are very close to those of the full model that suggests the model selection criteria are still useful for prediction purposes in the sense that they at least pick up a model close to the true model instead of picking a model close to the null model.

4. When the true model is full

   Again, all the criteria usually achieve the same level of performance in predictive loss as the non-selection method except CML works poorly:

it has larger predictive losses sometimes and can not identify the true model at all. Also, BIC and $MCML$ often works slightly worse than the others.

In conclusion, under the forward selection procedure the overall performance for Poisson models with $\rho = 0$ is:

$$CFB, FB.r, FB, HCML > BIC > AIC, CML, MCML$$

which agrees with findings in section 7.3 in that the FB criteria outperform others.

It is very important for us to realize that the performance results of the selection criteria for FB do not hold for other selection procedures. The reason is that if we do not put restrictions on k and ω, the criteria using Bayesian approaches tend to choose the full model when they face choices that consist of the full model and models close to the full model . FP starts searching from the null model and often has little chance to pass through to the full model unless the true model is close to full, hence the full model is not a big threat and FB and $FB.r$ performs equally well. This is also true for the stepwise procedure that starts searching from the null model. If we conduct backward selection procedure (BP), the criteria without restriction will very likely be distracted by the existence of the full model and stop at the full model. Figure (7.15) as well as Table (7.35)and (7.36) presents the results for Poisson models under BP when $n = 200$, $p = 20$, $\rho = 0$ and they confirm this: CML, $HCML$, FB and CFB have very close predictive losses to those of the non-selection method and cannot identify the right model at all with any number of nonzero components. $FB.r$ still has the best performance except when the true model is full. Even when the true model is full, $FB.r$ has similar predictive losses as the nonselection method though it cannot identify the right model. Overall, I can summarize the results for Poisson models with $\rho = 0$ under BP as:

$$FB.restrict > BIC > AIC > CML, HCML, FB, CFB$$

In addition, $MCML$ has comparable performance to BIC when the true model is not null but much worse performance when the true model is null.

Table 7.31: Average Predictive Loss Under Forward Procedure(FP)

Poisson, $\rho = 0$

p	n	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
10	100	5.34	5.27	5.29	5.29	5.51	6.01	6.59	5.94	7.28
		0.161	*0.158*	*0.159*	*0.156*	*0.158*	*0.144*	*0.175*	*0.145*	*0.147*
10	200	4.22	4.23	4.23	4.42	4.78	6.61	7.41	7.11	8.74
		0.099	*0.100*	*0.100*	*0.108*	*0.116*	*0.150*	*0.172*	*0.194*	*0.188*
10	500	3.23	3.25	3.26	3.35	3.35	4.77	3.94	4.74	6.08
		0.079	*0.079*	*0.079*	*0.082*	*0.079*	*0.094*	*0.084*	*0.109*	*0.106*
20	100	10.29	10.38	10.25	10.33	12.24	17.50	12.08	20.75	27.42
		0.268	*0.267*	*0.266*	*0.258*	*0.280*	*0.412*	*0.248*	*0.627*	*0.710*
20	200	8.62	8.63	8.66	8.75	8.91	10.39	9.33	10.89	12.23
		0.232	*0.231*	*0.231*	*0.228*	*0.217*	*0.184*	*0.218*	*0.207*	*0.192*
20	500	6.34	6.45	6.38	6.30	7.95	9.81	7.57	11.25	12.72
		0.129	*0.131*	*0.129*	*0.121*	*0.156*	*0.153*	*0.116*	*0.234*	*0.211*
50	200	29.95	29.66	29.64	27.65	33.02	33.93	28.59	37.80	47.27
		0.495	*0.490*	*0.490*	*0.448*	*0.502*	*0.403*	*0.435*	*0.617*	*0.555*

Table 7.32: Percent of Hits Under FP

Poisson, $\rho = 0$

p	n	cfb	fb.r	fb	hcml	bic	aic	cml	mcml
10	100	0.70	0.70	0.70	0.69	0.61	0.38	0.35	0.51
		0.012	*0.012*	*0.012*	*0.012*	*0.013*	*0.013*	*0.012*	*0.013*
10	200	0.73	0.73	0.73	0.72	0.66	0.39	0.39	0.54
		0.011	*0.011*	*0.011*	*0.012*	*0.012*	*0.013*	*0.013*	*0.013*
10	500	0.82	0.81	0.81	0.81	0.79	0.38	0.58	0.63
		0.010	*0.010*	*0.010*	*0.010*	*0.011*	*0.013*	*0.013*	*0.012*
20	100	0.54	0.44	0.55	0.55	0.29	0.06	0.23	0.16
		0.013	*0.013*	*0.013*	*0.013*	*0.012*	*0.006*	*0.011*	*0.009*
20	200	0.53	0.52	0.51	0.42	0.51	0.15	0.24	0.36
		0.013	*0.013*	*0.013*	*0.013*	*0.013*	*0.009*	*0.011*	*0.012*
20	500	0.80	0.73	0.79	0.76	0.58	0.21	0.43	0.42
		0.010	*0.011*	*0.011*	*0.011*	*0.013*	*0.010*	*0.013*	*0.013*
50	200	0.31	0.26	0.31	0.33	0.13	0.00	0.16	0.05
		0.009	*0.008*	*0.009*	*0.009*	*0.006*	*0.001*	*0.007*	*0.004*

Table 7.33: Average Predictive Loss by Nonzero Components Under FP Poisson, $\rho = 0, n = 200, p = 20$

q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
0	0.63	0.66	0.66	0.92	3.08	8.95	4.36	14.96	15.22
	0.05	*0.06*	*0.06*	*0.12*	*0.22*	*0.29*	*0.14*	*0.32*	*0.30*
4	2.77	2.78	2.78	2.93	3.28	5.38	2.91	3.46	8.42
	0.16	*0.16*	*0.16*	*0.17*	*0.19*	*0.23*	*0.17*	*0.19*	*0.25*
8	5.31	5.34	5.34	5.58	6.22	8.68	5.52	6.51	11.55
	0.17	*0.17*	*0.17*	*0.18*	*0.19*	*0.23*	*0.17*	*0.19*	*0.28*
12	13.70	14.02	14.02	15.62	12.91	16.37	15.78	13.67	20.28
	0.55	*0.56*	*0.56*	*0.61*	*0.53*	*0.59*	*0.61*	*0.54*	*0.69*
16	6.03	5.96	6.11	6.35	5.47	5.83	6.33	5.61	6.38
	0.17	*0.17*	*0.17*	*0.17*	*0.17*	*0.16*	*0.17*	*0.17*	*0.17*
20	23.30	23.03	23.03	21.11	22.51	17.11	21.09	21.12	11.50
	0.34	*0.34*	*0.34*	*0.38*	*0.35*	*0.35*	*0.39*	*0.35*	*0.31*

Table 7.34: No. of Hits by Nonzero Components Under FP Poisson, $\rho = 0, n = 200, p = 20$

q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
0	250	249	249	242	151	8	0	0	0
4	236	235	235	221	181	22	224	160	0
8	0	0	0	0	0	0	0	0	0
12	189	180	180	122	218	71	115	182	0
16	113	116	96	27	218	123	15	191	0
20	0	0	0	16	0	0	0	0	250

Table 7.35: Average Predictive Loss by Nonzero Components Under BP Poisson, $\rho = 0, n = 200, p = 20$

q	fb.r	bic	mcml	aic	cfb	fb	cml	hcml	ns
0	1.57	5.57	27.20	17.80	30.44	30.44	28.95	30.44	30.44
	0.253	*0.406*	*0.820*	*0.624*	*0.661*	*0.661*	*0.727*	*0.661*	*0.661*
4	4.01	5.50	6.06	11.31	18.06	18.06	18.07	18.06	18.06
	0.227	*0.305*	*0.319*	*0.451*	*0.517*	*0.517*	*0.517*	*0.517*	*0.517*
8	4.21	4.29	4.73	6.27	8.39	8.39	8.39	8.39	8.39
	0.184	*0.183*	*0.196*	*0.230*	*0.228*	*0.228*	*0.228*	*0.228*	*0.228*
12	16.42	17.65	15.21	15.26	17.60	17.60	17.55	17.60	17.60
	0.627	*0.672*	*0.582*	*0.554*	*0.512*	*0.512*	*0.512*	*0.512*	*0.512*
16	13.28	15.38	13.97	13.01	12.78	12.78	12.75	12.78	12.78
	0.386	*0.390*	*0.385*	*0.388*	*0.354*	*0.354*	*0.355*	*0.354*	*0.354*
20	9.77	9.53	9.57	10.56	12.93	12.93	12.89	12.93	12.93
	0.345	*0.372*	*0.329*	*0.337*	*0.348*	*0.348*	*0.346*	*0.348*	*0.348*

Table 7.36: No. of Hits by Nonzero Components Under BP Poisson, $\rho = 0, n = 200, p = 20$

q	fb.r	bic	mcml	aic	cfb	fb	cml	hcml
0	247	155	0	7	0	0	0	0
4	232	172	147	19	0	0	0	0
8	193	181	141	31	0	0	0	0
12	80	75	92	53	0	0	0	0
16	43	11	33	55	0	0	0	0
20	0	0	0	0	250	250	0	250

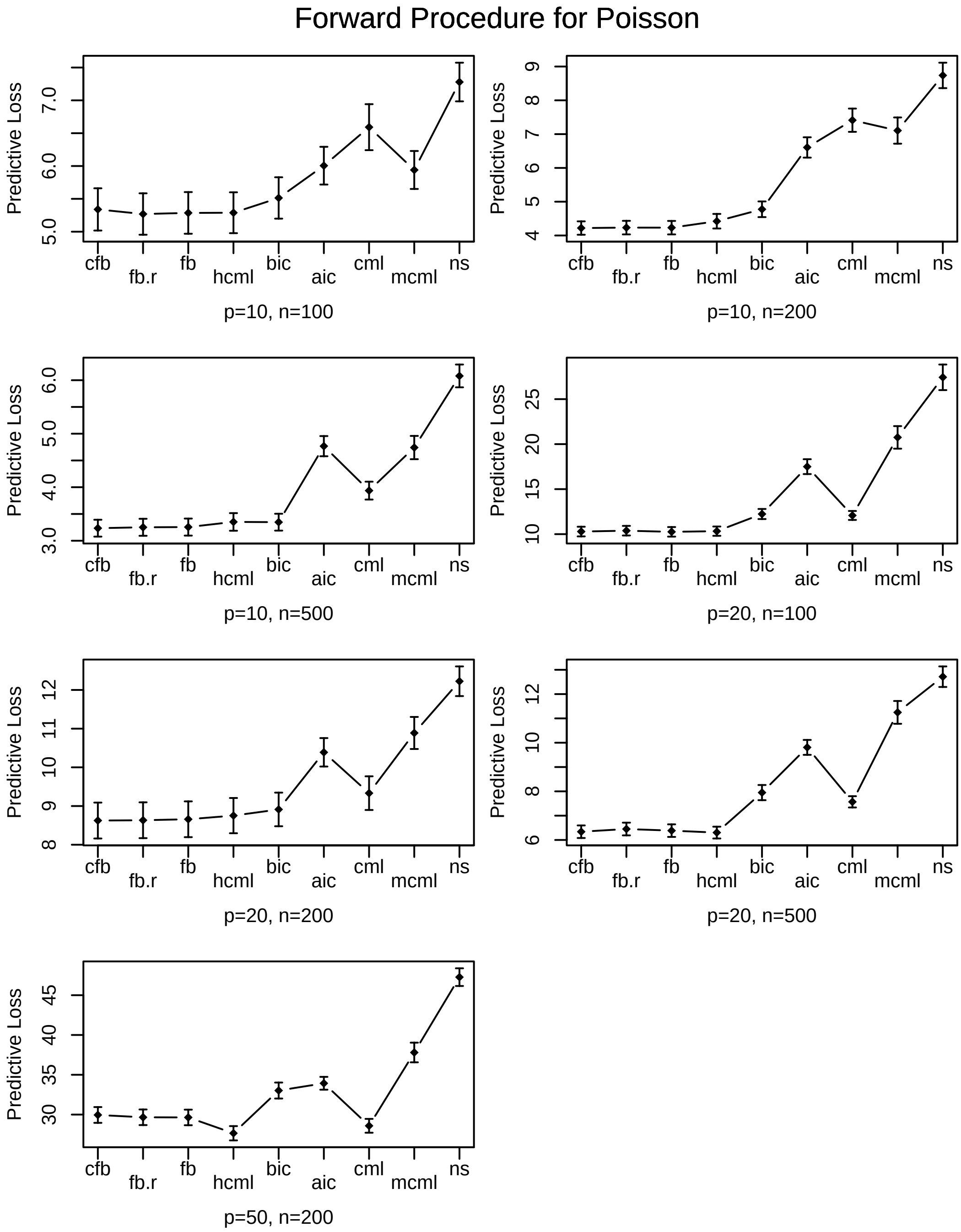

Figure 7.13: Average Predictive Loss With 95% CI Under FP

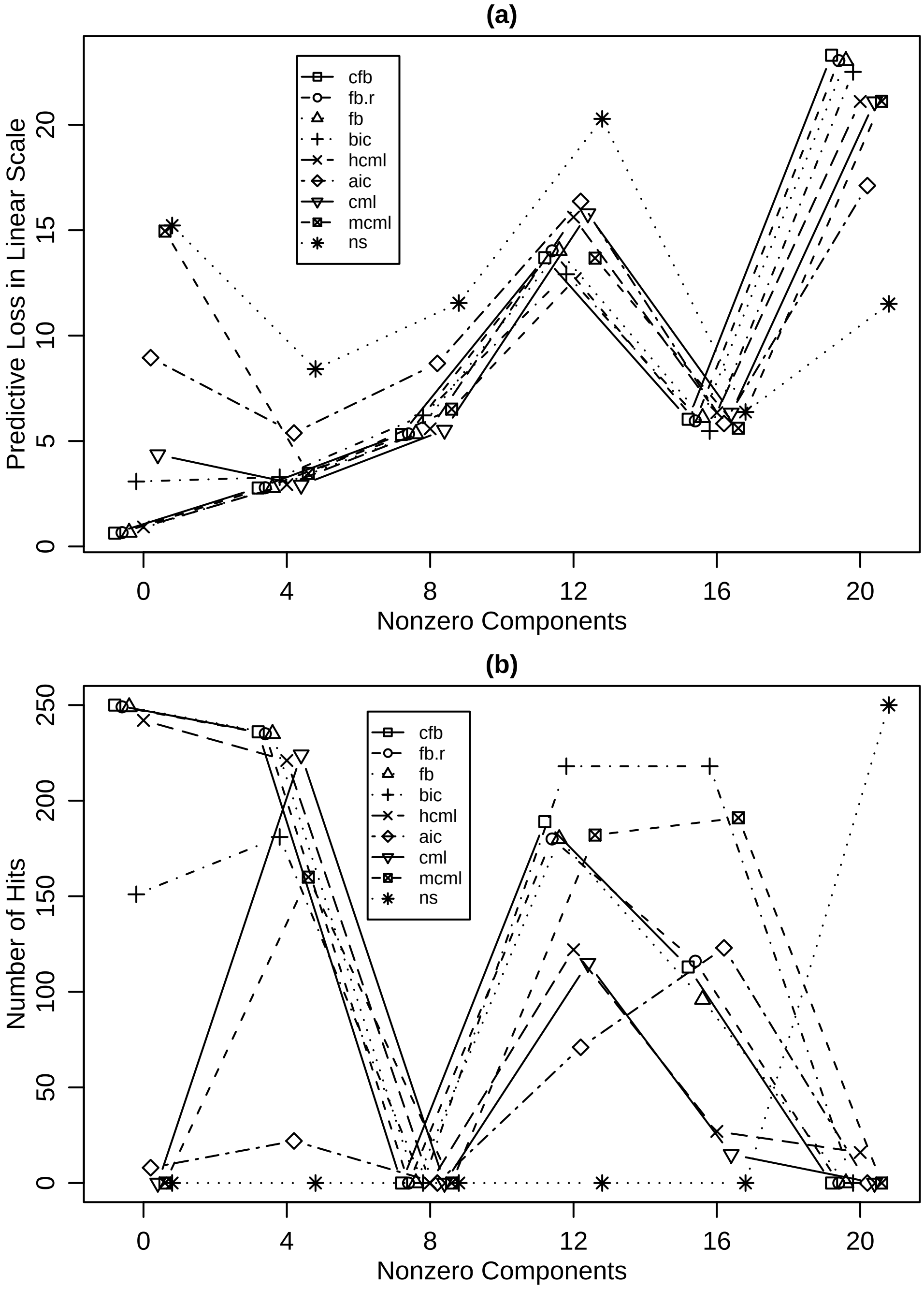

Figure 7.14: Comparison Under FP: Poisson, $\rho = 0, n = 200, p = 20$

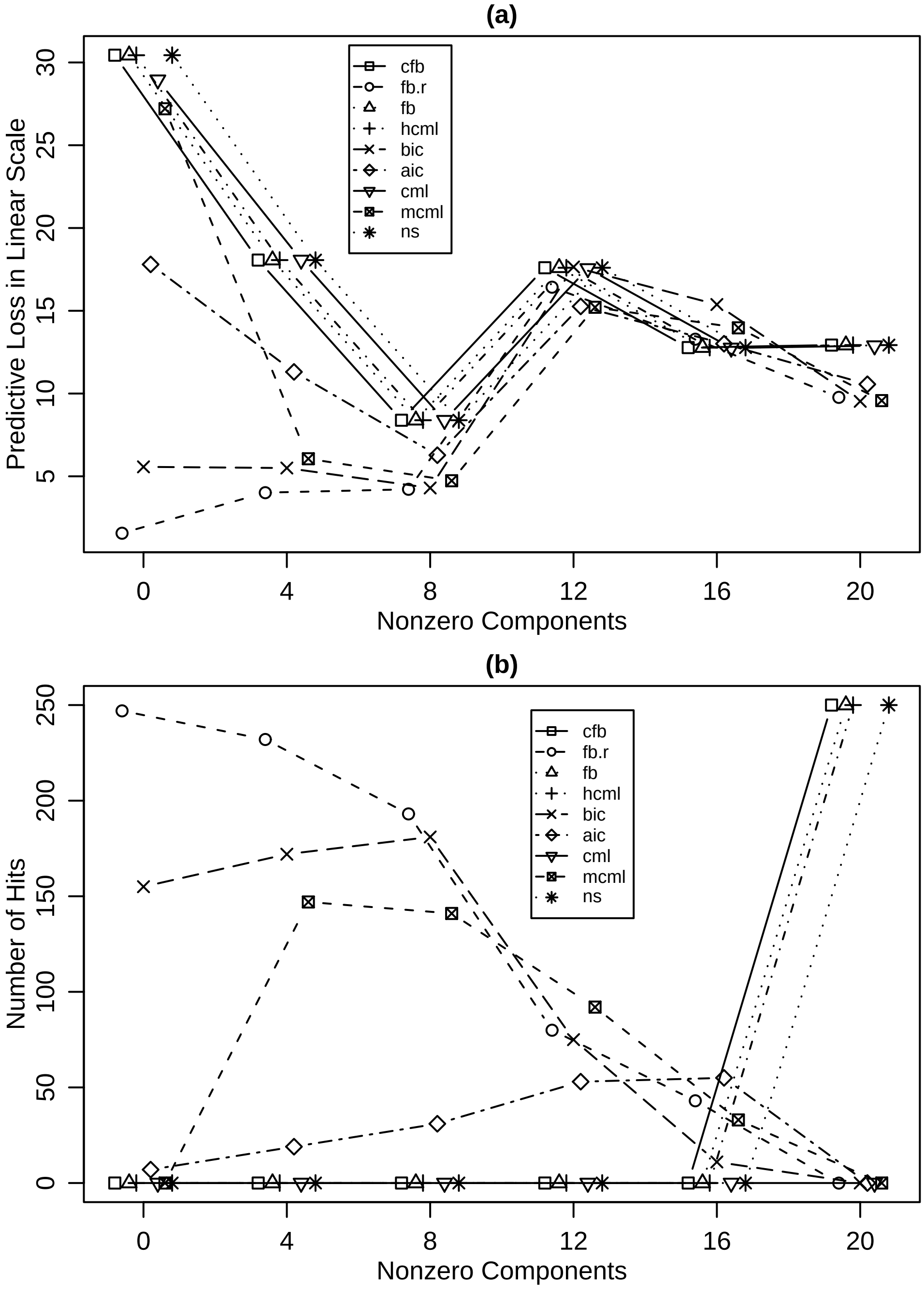

Figure 7.15: Comparison Under BP: Poisson, $\rho = 0, n = 200, p = 20$

7.7 Summary

In this chapter, the performance of selection criteria proposed in the dissertation is evaluated and compared with the classical criteria for GLM through the simulation results. There are several insightful findings here.

First, though the criteria under the three distributions (Poisson, Bernoulli and Normal) have different performance, we have seen that the FB criteria are consistently better than the EB criteria. This agrees with Berger (1985)[p.195], which comments that "it appears that the hierarchical Bayes approach is the superior methodology for general application".

Second, for all the three kinds of models I considered in the simulation, the two FB criteria, CFB and $FB.r$, usually have better overall performance than the classical ones, AIC and BIC. More attractively, for Poisson and Normal models, we have observed that these two FB criteria usually have the best performance at any number of nonzero components. This is really valuable in practice, which brings better performance in variable selection for sure for Poisson models and linear regression no matter where the true model is. For Logistic regression, the two FB criteria should achieve better performance in a long run.

Third, the performance of the EB criteria is a bit disappointing, especially for Logistic regression. However, $HCML$ works much better than AIC and sometimes better than BIC at hitting right models under Poisson and Normal models. We recall $HCML$ only applies the EB idea to hyperparameter c and applies the FB idea to ω. It is interesting to see this small application of the FB thought might be the reason of improving count of hits, as compared to CML and $MCML$.

All the evidences from simulation indicate that $FB.r$ or CFB should be applied for the best performance. It is helpful to discuss the two criteria from a calculational perspective. Conventionally, people think the Bayesian criteria are calculation-intensive. However, this is not true. $FB.r$ requires one-dimension numerical integration over a closed interval [0.5, 1]. Unlike a high dimensional integration, this is easily to be done. CFB only needs to evaluate a CDF of Gamma distribution. Therefore, they are not computationally expensive, especially under our modern computing environment. The seemingly complicated formula of the two criteria should not keep them from wide future use in variable selection, as they have excellent performance.

Before I finish this chapter, I suggest several methods of reducing the model space. It is impossible to go through all the 2^p GLM models when p is not small. One solution is to select a subset of models using the stepwise method then apply a selection criterion to the subset, like what I did in the simulation. An alternative is to use a selection procedure. As discussed in section 7.6, the backward procedure is not good for the EB or FB criteria except for $FB.r$. A forward or stepwise procedure that starts searching from the null model might be good in practice. Some other techniques are also available, like Occam's windows (Madigan and Raftery, 1994), stochastic search (George and McCulloch, 1993 and 1996; George, McCulloch and Tsay, 1994), etc.

Chapter 8

Discussion and Future Research

8.1 Discussion

In this dissertation I have presented a comprehensive hierarchical Bayesian solution to the variable selection problem of GLM. The two different Bayesian approaches, Fully Bayes and Empirical Bayes, have been explored after I achieved analytical tractability for GLM by proposing an Integrated Laplace Approximation. Different selection criteria were developed under each approach: CML, $MCML$ and $HCML$ are the EB criteria; FB, $FB.r$ and CFB are the FB criteria under noninformative hyperpriors. The original criterion FB that was developed directly under the FB framework tends to put a high probability on the full model. Hence, $FB.r$ was aimed to improve FB based on a method of restricting the integration region of hyperparameters, that avoids disturbance of the full model in variable selection. Also, CFB was proposed for improving the selection performance under FB based on a relaxation of hyperparameters to model-specific ones. I have shown these two criteria produce better performance than classical criteria AIC and BIC, besides achieving their original goals. Especially for Poisson and normal linear models, they usually have the best performance no matter how many variables are in the true model, which is valuable in practice. Another important finding is that the FB methods are consistently superior to the EB methods in selection performance.

Though major theoretical work has been done for GLM with a canonical link, I generalized all the work to GLM with a noncanonical link in chapter 6. I also considered normal linear models specifically to build a connection with previous research work.

8.2 Future Research

As direct extensions of the work I describe in this dissertation, I would like to pursue several areas of research.

8.2.1 Choosing The Knob Function for CFB

The CFB criterion with the knob function $\tau = \frac{1}{q+1}$ has demonstrated its excellent performance in variable selection. I chose the above knob function based on some graphical information of the data-independent penalty part in the CFB posterior as well as its simple form. By no means can we believe that this is the best knob function we can get. For different distributions, better performance may be achieved by using different knob functions. Especially for Logistic regression, both $FB.r$ and CFB do not have uniformly better performance than classical criteria at any number of nonzero components, though they win in overall performance. Since CFB can be made very flexible by choosing knob functions, one natural question is whether we get a CFB criterion that is uniformly better than classical ones. There should be a lot of space to explore in finding suitable knob functions.

One would feel more comfortable with CFB if we can develop good theoretical guidelines on choosing a proper knob function. In this dissertation, I have not put much effort into it. I would like to do this in the near future.

8.2.2 Incorporating Expert Knowledge

A major strength of a Bayesian approach is that it allows for incorporation of expert opinion along with information from the data. In this dissertation I adopted a set of prior and hyperprior distributions with the aim of reflecting the priors of a person with little prior information. While this natural approach to prior selection was successful, Spiegelhalter et al. (1993) and Lauritzen et al. (1994) analyzed the benefits of incorporating informative prior distributions and demonstrated improved predictive performance with informative priors. This suggests that we could achieve better results if expert knowledge that may be available is considered. I would like to consider how we can formulate this expert knowledge mathematically and incorporate it into the framework I have built here. Usually, expert knowledge is verbal and sometimes vague though informative so we have to translate it into mathematical language. For example, if the expert knowledge tells us that some variables must be in the model, how do we formulate this information? The second example is, if the expert knowledge informs us that some variables are very likely to appear in the model and some others are not, how can we incorporate this into the framework? Obviously, it is harder than the previous example. It is also hard to incorporate expert knowledge into hyperpriors since hyperparameters are usually not meaningful in practice. For example, if we know that the underlying model might only have a few variables, how do we adjust our hyperpriors? I would like to see how Bayesian variable selection works under expert knowledge.

8.2.3 Accounting for Model Uncertainty

In the common practice of data analysis, people select a single "best" model based on some criteria and then make inferences as if the selected model were the true model. This has been criticized for disregarding model uncertainty that leads to over-confident inferences and decisions. Unlike AIC and BIC, the criteria developed here are actually posterior of models, which can easily account for model uncertainty. Hence, in a subsequent inference, we might do Bayesian model averaging (BMA) based on the model posterior of CFB or $FB.r$ and expect an improvement in predictive performance. BMA has been a active research field for a while (Madigan, Raftery, Volinsky and Hoeting, 1996; Raftery, Madigan and Volinsky, 1996; Raftery, Madigan and Hoeting, 1997; Hoeting, Madigan and Raftery, 1999; Clyde, 1999). I would like to implement BMA under the hierarchical framework in this dissertation and consider practical matters.

8.2.4 Other Future Research

There are some other potential research topics I am interested in.

In this work, a model with the largest posterior is chosen as the final model. However, this is a pretty arbitrary decision. For example, Model 1 has the largest posterior probability 0.10 and model 2 that ranks second has posterior probability 0.09. In current practice, Model 1 is chosen and Model 2 is thrown away, which does not make much sense as Model 2 is also likely to be the true model. It would be more reasonable that we use some test data that is not used in calculating model posterior, and test the candidates' predictive performance, finally choose the one with the best performance. This simple idea might bring better odds that the true model is chosen. Are there any other better methods to make a final decision?

Also, can we combine the strength of different criteria? Certainly different criteria can indicate different final models. We have observed that for Logistic regression, no criterion is uniformly better than others. Some criteria work better under some situations and worse under others. If we can combine their strength, better performance can be achieved in variable selection.

Appendices

Appendix A

Calculation of $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)$

Let's calculate $\tilde{p}\left(\mathbf{Y}|\gamma,c\right)=\int_{\mathbf{R}^{q_\gamma+1}}\tilde{p}\left(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma\right)p\left(\boldsymbol{\beta}_\gamma|\gamma,c\right)d\boldsymbol{\beta}_\gamma$ where the prior of $\boldsymbol{\beta}_\gamma$ is

$$\boldsymbol{\beta}_\gamma|\gamma,c\sim\mathbf{N}_{q_\gamma+1}(\mathbf{m}_\gamma,\ \mathbf{U})$$

$$\begin{aligned}
&\tilde{p}\left(\mathbf{Y}|\boldsymbol{\beta}_\gamma,\gamma\right)p\left(\boldsymbol{\beta}_\gamma|\gamma,c\right)\\
&=\ (2\pi)^{-\frac{q_\gamma+1}{2}}|\mathbf{U}|^{-\frac{1}{2}}\exp\left\{\frac{1}{\phi}\left[\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-\mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}-\right.\right.\\
&\quad\left.-\frac{1}{2}(\boldsymbol{\beta}_\gamma-\hat{\boldsymbol{\beta}}_\gamma)^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma(\boldsymbol{\beta}_\gamma-\hat{\boldsymbol{\beta}}_\gamma)-\frac{1}{2}(\boldsymbol{\beta}_\gamma-\mathbf{m}_\gamma)^T\left(\frac{\mathbf{U}}{\phi}\right)^{-1}(\boldsymbol{\beta}_\gamma-\mathbf{m}_\gamma)\right]\\
&\quad\left.+\mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right\}\\
&=\ (2\pi)^{-\frac{q_\gamma+1}{2}}|\mathbf{U}|^{-\frac{1}{2}}\exp\left\{\frac{1}{2\phi}\left[-\boldsymbol{\beta}_\gamma^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma+\left(\frac{\mathbf{U}}{\phi}\right)^{-1})\boldsymbol{\beta}_\gamma\right.\right.\\
&\quad+2\boldsymbol{\beta}_\gamma^T\left(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma+\left(\frac{\mathbf{U}}{\phi}\right)^{-1}\mathbf{m}_\gamma\right)+2\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-2\mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}\\
&\quad\left.\left.-\hat{\boldsymbol{\beta}}_\gamma^T\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma-\mathbf{m}_\gamma^T\left(\frac{\mathbf{U}}{\phi}\right)^{-1}\mathbf{m}_\gamma+2\phi\mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right]\right\}
\end{aligned}$$

Recall the formula for completing the square

$$\mathbf{Z}^TA\mathbf{Z}+\mathbf{Z}^TB+C=\left(\mathbf{Z}+\frac{1}{2}A^{-1}B\right)^TA\left(\mathbf{Z}+\frac{1}{2}A^{-1}B\right)+\left(C-\frac{1}{4}B^TA^{-1}B\right)$$

In this case, treating $\boldsymbol{\beta}_\gamma$ as $\mathbf{Z}$, we have

$$
\begin{aligned}
A &= \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma + \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \\
B &= 2\left[-\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \mathbf{m}_\gamma\right] \\
C &= -2\mathbf{Y}^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma + 2\mathbf{b}^T(\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma) \cdot \mathbf{1} + \hat{\boldsymbol{\beta}}_\gamma^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma + \mathbf{m}_\gamma^T \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \mathbf{m}_\gamma \\
&\quad -2\phi \mathbf{c}^T(\mathbf{Y}, \phi) \cdot \mathbf{1}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) p\,(\boldsymbol{\beta}_\gamma|\gamma, c) \\
&= (2\pi)^{-\frac{q_\gamma+1}{2}} |\mathbf{U}|^{-\frac{1}{2}} \exp\left\{-\frac{(\boldsymbol{\beta}_\gamma + \frac{1}{2}A^{-1}B)^T A(\boldsymbol{\beta}_\gamma + \frac{1}{2}A^{-1}B) + (C - \frac{1}{4}B^T A^{-1} B)}{2\phi}\right\}
\end{aligned}
$$

Therefore,

$$
\tilde{p}\,(\mathbf{Y}|\gamma, c) = \int_{-\infty}^{+\infty} \tilde{p}\,(\mathbf{Y}|\boldsymbol{\beta}_\gamma, \gamma) p\,(\boldsymbol{\beta}_\gamma|\gamma) d\boldsymbol{\beta}_\gamma
$$

$$
\begin{aligned}
&= |\mathbf{U}|^{-\frac{1}{2}} \overbrace{\int_{-\infty}^{+\infty} (2\pi)^{-\frac{q_\gamma+1}{2}} \left|\frac{A}{\phi}\right|^{\frac{1}{2}} \exp\left\{-\frac{(\boldsymbol{\beta}_\gamma + \frac{1}{2}A^{-1}B)^T A(\boldsymbol{\beta}_\gamma + \frac{1}{2}A^{-1}B)}{2\phi}\right\} d\boldsymbol{\beta}_\gamma}^{=1} \\
&\quad \left|\phi A^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{C - \frac{1}{4}B^T A^{-1} B}{2\phi}\right\} \\
&= |\mathbf{U}|^{-\frac{1}{2}} \left|\phi A^{-1}\right|^{\frac{1}{2}} \exp\left\{-\frac{C - \frac{1}{4}B^T A^{-1} B}{2\phi}\right\}
\end{aligned}
$$

Now set

$$
\left(\frac{\mathbf{U}}{\phi}\right)^{-1} = \frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{c}
$$

then

$$A = \left(1 + \frac{1}{c}\right) \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma$$

Thus,

$$\tilde{p}\left(\mathbf{Y}|\gamma, c\right) = (c+1)^{-\frac{q_\gamma+1}{2}} \exp\left\{-\frac{C - \frac{1}{4}B^T A^{-1} B}{2\phi}\right\}$$

Now let's look at $\frac{1}{4}B^T A^{-1} B$. Since

$$\left(\frac{\mathbf{U}}{\phi}\right)^{-1} = \frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{c}$$

We have

$$\begin{aligned}
\frac{1}{4}B^T A^{-1} B = & \left(-\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \mathbf{m}_\gamma\right)^T \frac{c}{c+1} (\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)^{-1} \\
& \cdot \left(-\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \mathbf{m}_\gamma\right) \\
= & \left(-\hat{\boldsymbol{\beta}}_\gamma^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma - \mathbf{m}_\gamma^T \frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{c}\right) \frac{c}{c+1} (\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)^{-1} \\
& \cdot \left(-\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \frac{\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma}{c} \mathbf{m}_\gamma\right) \\
= & \frac{c}{c+1} \left(\hat{\boldsymbol{\beta}}_\gamma + \frac{\mathbf{m}_\gamma}{c}\right)^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \left(\hat{\boldsymbol{\beta}}_\gamma + \frac{\mathbf{m}_\gamma}{c}\right)
\end{aligned}$$

$$\begin{aligned}
& C - \frac{1}{4}B^T A^{-1} B \\
= & -2\mathbf{Y}^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma + 2\mathbf{b}^T (\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma) \cdot \mathbf{1} + \hat{\boldsymbol{\beta}}_\gamma^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma + \mathbf{m}_\gamma^T \left(\frac{\mathbf{U}}{\phi}\right)^{-1} \mathbf{m}_\gamma \\
& -2\phi \mathbf{c}^T (\mathbf{Y}, \phi) \cdot \mathbf{1} - \frac{c}{c+1} \left(\hat{\boldsymbol{\beta}}_\gamma + \frac{\mathbf{m}_\gamma}{c}\right)^T \mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma \left(\hat{\boldsymbol{\beta}}_\gamma + \frac{\mathbf{m}_\gamma}{c}\right) \\
= & -2\mathbf{Y}^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma + 2\mathbf{b}^T (\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma) \cdot \mathbf{1} - 2\phi \mathbf{c}^T (\mathbf{Y}, \phi) \cdot \mathbf{1} \\
& + \frac{1}{c+1} (\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T (\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)
\end{aligned}$$

Thus,

$$\begin{aligned}
\tilde{p}\,(\mathbf{Y}|\gamma, c) &= (c+1)^{-\frac{q_\gamma+1}{2}} \exp\left\{-\frac{C - \frac{1}{4}B^T A^{-1} B}{2\phi}\right\} \\
&= (c+1)^{-\frac{q_\gamma+1}{2}} \exp\left\{\frac{\mathbf{Y}^T \mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma \hat{\boldsymbol{\beta}}_\gamma) \cdot \mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y}, \phi) \cdot \mathbf{1}\right\} \\
&\quad \cdot \exp\left\{-\frac{1}{2(c+1)}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T \frac{(\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)}{\phi}(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)\right\}
\end{aligned}$$

when the prior of $\boldsymbol{\beta}_\gamma$ is

$$p(\boldsymbol{\beta}_\gamma|\gamma, c) \sim \mathbf{N}_{q_\gamma+1}(\mathbf{m}_\gamma,\ c\phi(\mathbf{X}_\gamma^T \mathbf{V}_\gamma \mathbf{X}_\gamma)^{-1})$$

Appendix B

Calculation of $\pi\left(\gamma|\mathbf{Y}\right)$ Based on Restricted Integration Region

$$p(\mathbf{Y}|\gamma, c)\,\pi(\gamma|\omega) = \omega^{q_\gamma}(1-\omega)^{p-q_\gamma}k^{\frac{q_\gamma+1}{2}}\exp\left[-\frac{T_\gamma}{2}k\right]\cdot\hat{L}_\gamma\cdot(1+O(n^{-1}))$$

where

$$\hat{L}_\gamma = \exp\left[\frac{\mathbf{Y}^T\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma - \mathbf{b}^T(\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma)\cdot\mathbf{1}}{\phi} + \mathbf{c}^T(\mathbf{Y},\phi)\cdot\mathbf{1}\right]$$

and

$$T_\gamma = \frac{(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)^T(\mathbf{X}_\gamma^T\mathbf{V}_\gamma\mathbf{X}_\gamma)(\hat{\boldsymbol{\beta}}_\gamma - \mathbf{m}_\gamma)}{\phi}$$

Now let's consider the conjugate priors on both ω and k, i.e.,

$$\begin{aligned}\omega &\sim Beta(\alpha,\beta)\\ k &\sim Truncated\ Gamma(a,b), k\in(0,1)\end{aligned}$$

Therefore,

$$\begin{aligned}&\pi\left(\gamma|\mathbf{Y}\right)\propto\\ &\quad\hat{L}_\gamma\iint_D\omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1}k^{\frac{q_\gamma+1}{2}+a-1}\exp\left[-\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)k\right]d\omega dk\\ &\quad\cdot(1+O(n^{-1}))\end{aligned}$$

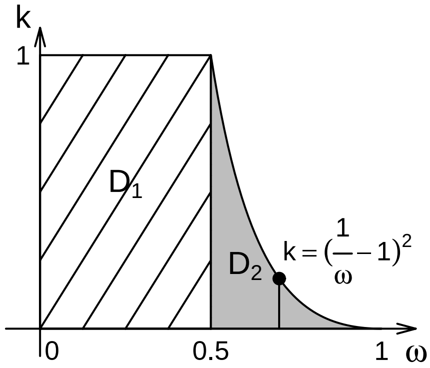

Figure B.1: Integration Area

where $\{D : \omega, k | 2\log\frac{1-\omega}{\omega} - \log k \geq 0\}$. D can be decomposed to D_1 and D_2 as shown below.

There are two cases in the integration part depending on whether $\frac{T_\gamma}{2} + \frac{1}{b}$ equals zero or not. The two cases are discussed separately.

First, let's look at case 1: $\frac{T_\gamma}{2} + \frac{1}{b} > 0$. Then,

$$
\begin{aligned}
&\iint_{D_1} \omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1} k^{\frac{q_\gamma+1}{2}+a-1} \exp\left[-\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)k\right] d\omega dk \\
= &\int_0^{0.5} \omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1} d\omega \int_0^1 k^{\frac{q_\gamma+1}{2}+a-1} \exp\left[-\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)k\right] dk \\
= &\frac{\Gamma(q_\gamma+\alpha)\Gamma(p-q_\gamma+\beta)}{\Gamma(p+\alpha+\beta)} B_{(q_\gamma+\alpha, p-q_\gamma+\beta)}(0.5) \\
&\cdot\Gamma\left(\frac{q_\gamma+1}{2}+a\right) \cdot \left(\frac{T_\gamma}{2}+\frac{1}{b}\right)^{-\frac{q_\gamma+1}{2}-a} \cdot G_{(\frac{q_\gamma+1}{2}+a,1)}\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)
\end{aligned}
$$

$$\begin{aligned}
&\iint_{D_2} \omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1}k^{\frac{q_\gamma+1}{2}+a-1}\exp\left[-\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)k\right]d\omega dk \\
&= \int_{0.5}^{1}\omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1}d\omega\int_0^{(\frac{1}{\omega}-1)^2}k^{\frac{q_\gamma+1}{2}+a-1}\exp\left[-\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)k\right]dk \\
&= \Gamma\left(\frac{q_\gamma+1}{2}+a\right)\cdot\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)^{-\frac{q_\gamma+1}{2}-a}\int_{0.5}^{1}\omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1} \\
&\quad \cdot G_{(\frac{q_\gamma+1}{2}+a,1)}\left(\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)\left(\frac{1}{\omega}-1\right)^2\right)d\omega
\end{aligned}$$

Thus,

$$\begin{aligned}
&\pi\left(\gamma|\mathbf{Y}\right) \propto \hat{L}_\gamma\Gamma\left(\frac{q_\gamma+1}{2}+a\right)\cdot\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)^{-\frac{q_\gamma+1}{2}-a} \\
&\quad \cdot\left\{\frac{\Gamma(q_\gamma+\alpha)\Gamma(p-q_\gamma+\beta)}{\Gamma(p+\alpha+\beta)}B_{(q_\gamma+\alpha,p-q_\gamma+\beta)}(0.5)\cdot G_{(\frac{q_\gamma+1}{2}+a,1)}\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)\right. \\
&\quad \left.+\int_{0.5}^{1}\omega^{q_\gamma+\alpha-1}(1-\omega)^{p-q_\gamma+\beta-1}\cdot G_{(\frac{q_\gamma+1}{2}+a,1)}\left(\left(\frac{T_\gamma}{2}+\frac{1}{b}\right)\left(\frac{1}{\omega}-1\right)^2\right)d\omega\right\} \\
&\quad (1+O(n^{-1}))
\end{aligned}$$

Under the uniform priors on k and ω and $T_\gamma \neq 0$, we have $\alpha = 1$, $\beta = 1$, $a = 1$, $b = +\infty$, and the above formula becomes (4.5).

Second, let's look at case 2: $\frac{T_\gamma}{2} + \frac{1}{b} = 0$. Since both T_γ and b are non-negative, it can only happen when $T_\gamma = 0$ and $b = \infty$, i.e., $\mathbf{m}_\gamma = \hat{\boldsymbol{\beta}}_\gamma$. We will focus on non-informative priors on both ω and k, i.e., $\alpha = 1, \beta = 1, a = 1, b = \infty$. Then,

$$\pi\left(\gamma|\mathbf{Y}\right) \propto \hat{L}_\gamma\iint_D \omega^{q_\gamma}(1-\omega)^{p-q_\gamma}k^{\frac{q_\gamma+1}{2}}d\omega dk(1+O(n^{-1}))$$

$$\begin{aligned}
&\iint_{D_1} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} k^{\frac{q_\gamma+1}{2}} d\omega dk \\
&= \int_{0.5}^{1} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} d\omega \int_0^1 k^{\frac{q_\gamma+1}{2}} dk \\
&= \frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)} B_{(q_\gamma+1,p-q_\gamma+1)}(0.5) \cdot \frac{2}{q_\gamma+3}
\end{aligned}$$

$$\begin{aligned}
&\iint_{D_2} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} k^{\frac{q_\gamma+1}{2}} d\omega dk \\
&= \int_{0.5}^{1} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} d\omega \int_0^{(\frac{1}{\omega}-1)^2} k^{\frac{q_\gamma+1}{2}} dk \\
&= \int_{0.5}^{1} \omega^{q_\gamma}(1-\omega)^{p-q_\gamma} \frac{2}{q_\gamma+3} \left(\frac{1}{\omega}-1\right)^{q_\gamma+3} d\omega \\
&= \frac{2}{q_\gamma+3} \int_{0.5}^{1} \omega^{-3}(1-\omega)^{p+3} d\omega
\end{aligned}$$

Thus,

$$\begin{aligned}
\pi(\gamma|\mathbf{Y}) \quad \propto \quad & \hat{L}_\gamma \frac{2}{q_\gamma+3} \left[\frac{\Gamma(q_\gamma+1)\Gamma(p-q_\gamma+1)}{\Gamma(p+2)} B_{(q_\gamma+1,p-q_\gamma+1)}(0.5) \right. \\
& \left. + \int_{0.5}^{1} \omega^{-3}(1-\omega)^{p+3} d\omega \right] (1+O(n^{-1}))
\end{aligned}$$

This is exactly formula (4.7).

Appendix C

Simulation Results By (d, n, p, ρ, q)

C.1 Poisson Models

Table C.1: Average Predictive Loss: Poisson, $p = 10, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	0.39	0.39	0.39	1.34	1.05	3.18	2.36	3.41	3.75	3.97
		0.04	*0.04*	*0.04*	*0.16*	*0.09*	*0.16*	*0.12*	*0.14*	*0.13*	*0.12*
100	2	1.55	1.73	1.73	1.73	2.17	1.83	3.30	1.80	2.23	5.03
		0.12	*0.13*	*0.13*	*0.13*	*0.14*	*0.13*	*0.17*	*0.13*	*0.14*	*0.18*
100	4	2.09	2.05	2.05	2.05	2.49	2.12	3.32	2.11	2.49	4.52
		0.10	*0.11*	*0.11*	*0.11*	*0.13*	*0.11*	*0.14*	*0.11*	*0.13*	*0.16*
100	6	4.60	6.36	6.37	6.55	6.35	7.20	6.77	6.93	6.41	7.71
		0.21	*0.23*	*0.22*	*0.24*	*0.22*	*0.27*	*0.27*	*0.25*	*0.23*	*0.30*
100	8	6.16	7.34	7.15	7.45	6.81	7.68	7.26	7.56	6.95	7.70
		0.31	*0.34*	*0.32*	*0.34*	*0.31*	*0.36*	*0.33*	*0.35*	*0.31*	*0.37*
100	10	4.63	6.28	6.73	5.68	8.35	4.84	5.95	5.65	7.88	4.63
		0.18	*0.25*	*0.24*	*0.24*	*0.23*	*0.19*	*0.20*	*0.20*	*0.24*	*0.18*
200	0	0.71	0.77	0.80	2.56	1.75	6.50	4.84	6.90	7.61	7.99
		0.06	*0.09*	*0.09*	*0.30*	*0.15*	*0.30*	*0.23*	*0.27*	*0.24*	*0.22*
200	2	2.44	2.66	2.66	2.66	3.09	2.78	5.56	2.73	3.21	8.01
		0.18	*0.19*	*0.19*	*0.19*	*0.21*	*0.20*	*0.25*	*0.20*	*0.22*	*0.25*
200	4	1.76	1.89	1.92	1.92	2.01	2.02	2.70	2.02	2.05	3.38
		0.09	*0.09*	*0.09*	*0.09*	*0.10*	*0.10*	*0.11*	*0.10*	*0.10*	*0.12*
200	6	2.41	2.18	2.23	2.23	2.27	2.30	2.99	2.31	2.30	3.80
		0.11	*0.11*	*0.11*	*0.11*	*0.11*	*0.11*	*0.13*	*0.11*	*0.11*	*0.13*
200	8	7.51	8.85	8.60	8.65	8.95	8.70	8.14	8.56	8.24	8.82
		0.34	*0.41*	*0.39*	*0.40*	*0.43*	*0.38*	*0.34*	*0.37*	*0.38*	*0.35*
200	10	3.24	3.89	4.03	3.79	4.95	3.40	3.79	3.91	4.70	3.24
		0.10	*0.12*	*0.12*	*0.12*	*0.14*	*0.11*	*0.11*	*0.11*	*0.14*	*0.10*
500	0	0.73	0.73	0.95	3.54	1.75	7.64	5.81	8.26	8.97	9.36
		0.07	*0.07*	*0.13*	*0.41*	*0.19*	*0.39*	*0.29*	*0.34*	*0.31*	*0.29*
500	2	2.26	2.40	2.40	2.40	2.74	2.43	5.16	2.41	2.70	7.58
		0.13	*0.14*	*0.14*	*0.14*	*0.16*	*0.14*	*0.22*	*0.14*	*0.16*	*0.24*
500	4	2.38	2.49	2.51	2.51	2.58	2.61	3.68	2.64	2.65	4.56
		0.12	*0.13*	*0.13*	*0.13*	*0.13*	*0.13*	*0.15*	*0.14*	*0.13*	*0.15*
500	6	4.56	4.90	4.90	4.91	4.71	4.99	5.44	5.03	4.89	6.25
		0.25	*0.26*	*0.26*	*0.26*	*0.25*	*0.26*	*0.27*	*0.27*	*0.26*	*0.27*
500	8	6.52	7.06	7.02	7.08	6.85	7.27	7.11	7.17	6.92	7.58
		0.29	*0.30*	*0.30*	*0.30*	*0.31*	*0.30*	*0.30*	*0.30*	*0.31*	*0.30*
500	10	8.54	8.56	8.56	8.54	9.43	8.54	8.56	21.23	8.89	8.54
		0.36	*0.36*	*0.36*	*0.36*	*0.44*	*0.36*	*0.36*	*0.38*	*0.40*	*0.36*

Table C.2: Average Predictive Loss: Poisson, $p = 20, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.56	1.56	1.56	8.70	8.30	31.89	20.99	31.78	36.96	37.57
		0.14	*0.14*	*0.14*	*1.40*	*0.57*	*1.54*	*0.92*	*1.49*	*1.23*	*1.20*
100	4	3.76	4.29	4.47	4.47	6.08	4.87	12.35	4.85	7.34	21.78
		0.25	*0.27*	*0.28*	*0.28*	*0.38*	*0.30*	*0.58*	*0.30*	*0.44*	*0.72*
100	8	5.35	7.06	7.11	7.11	7.20	10.01	10.01	9.21	8.02	15.37
		0.23	*0.37*	*0.36*	*0.36*	*0.36*	*0.56*	*0.41*	*0.54*	*0.37*	*0.50*
100	12	8.87	13.37	12.40	13.69	12.10	15.01	12.17	14.95	11.79	15.01
		0.33	*0.59*	*0.55*	*0.58*	*0.56*	*0.50*	*0.49*	*0.50*	*0.49*	*0.50*
100	16	25.08	8.37	8.69	8.69	11.18	10.00	20.47	9.51	14.13	35.39
		0.95	*0.50*	*0.53*	*0.53*	*0.68*	*0.69*	*0.95*	*0.61*	*0.82*	*1.31*
100	20	28.81	31.72	33.51	30.34	38.66	28.81	31.69	28.93	34.34	28.81
		1.13	*1.19*	*1.18*	*1.16*	*1.01*	*1.13*	*1.11*	*1.14*	*1.09*	*1.13*
200	0	1.12	1.12	1.28	4.44	4.05	17.77	13.28	18.72	22.42	22.84
		0.08	*0.08*	*0.14*	*0.68*	*0.35*	*0.85*	*0.52*	*0.77*	*0.55*	*0.52*
200	4	2.40	2.80	2.80	2.80	3.42	2.93	6.90	2.92	3.80	11.35
		0.13	*0.17*	*0.17*	*0.17*	*0.20*	*0.18*	*0.26*	*0.18*	*0.22*	*0.29*
200	8	7.09	4.81	4.84	4.84	5.98	5.14	12.18	5.08	6.77	20.23
		0.31	*0.33*	*0.33*	*0.33*	*0.39*	*0.35*	*0.52*	*0.35*	*0.42*	*0.62*
200	12	7.04	8.41	8.44	8.70	8.10	11.03	9.24	10.52	8.24	11.23
		0.24	*0.30*	*0.30*	*0.32*	*0.31*	*0.33*	*0.30*	*0.34*	*0.28*	*0.32*
200	16	7.40	10.44	10.35	10.34	10.72	9.69	9.90	9.80	10.48	9.65
		0.21	*0.26*	*0.25*	*0.26*	*0.25*	*0.26*	*0.26*	*0.27*	*0.26*	*0.26*
200	20	12.21	12.70	12.71	12.81	12.59	12.79	12.53	12.92	12.77	12.21
		0.34	*0.24*	*0.24*	*0.25*	*0.24*	*0.29*	*0.29*	*0.29*	*0.24*	*0.34*
500	0	1.34	1.34	1.42	4.14	3.29	18.53	14.79	21.02	25.55	26.11
		0.15	*0.15*	*0.17*	*0.72*	*0.35*	*1.02*	*0.56*	*0.88*	*0.61*	*0.57*
500	4	1.48	1.58	1.58	1.58	1.83	1.61	3.53	1.60	1.94	5.45
		0.07	*0.08*	*0.08*	*0.08*	*0.09*	*0.08*	*0.12*	*0.08*	*0.10*	*0.12*
500	8	1.99	2.72	2.67	2.67	2.60	2.67	3.53	2.68	2.61	4.64
		0.09	*0.14*	*0.14*	*0.14*	*0.12*	*0.12*	*0.11*	*0.12*	*0.12*	*0.11*
500	12	13.82	12.60	12.85	12.85	12.75	13.32	17.14	13.32	13.58	21.13
		0.44	*0.44*	*0.44*	*0.44*	*0.44*	*0.45*	*0.50*	*0.45*	*0.46*	*0.53*
500	16	7.63	8.63	8.55	8.73	7.87	9.39	8.68	9.28	8.02	9.43
		0.23	*0.25*	*0.25*	*0.25*	*0.25*	*0.25*	*0.25*	*0.25*	*0.25*	*0.25*
500	20	19.34	19.34	19.34	19.34	22.93	19.34	19.55	33.73	20.22	19.34
		0.49	*0.49*	*0.49*	*0.49*	*0.74*	*0.49*	*0.51*	*0.52*	*0.57*	*0.49*

Table C.3: Average Predictive Loss: Poisson, $p = 50, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
200	0	1.21	1.21	1.31	9.17	9.29	61.16	35.99	60.07	67.89	68.14
		0.10	*0.10*	*0.14*	*1.74*	*0.59*	*1.96*	*1.08*	*1.98*	*1.30*	*1.28*
200	5	2.34	2.63	2.65	2.65	4.62	2.91	12.67	2.89	5.33	25.73
		0.14	*0.15*	*0.15*	*0.15*	*0.22*	*0.16*	*0.38*	*0.16*	*0.24*	*0.48*
200	10	5.96	8.61	8.48	8.48	10.16	9.24	21.80	9.05	12.99	42.79
		0.21	*0.45*	*0.43*	*0.43*	*0.46*	*0.45*	*0.66*	*0.43*	*0.54*	*0.89*
200	15	6.68	7.83	7.97	7.97	8.52	12.12	16.34	10.01	11.45	28.05
		0.21	*0.30*	*0.30*	*0.30*	*0.32*	*0.69*	*0.47*	*0.52*	*0.44*	*0.56*
200	20	10.95	14.10	14.25	14.25	14.97	16.17	23.67	15.68	17.73	38.20
		0.25	*0.37*	*0.38*	*0.38*	*0.39*	*0.57*	*0.58*	*0.50*	*0.47*	*0.72*
200	25	15.74	42.81	42.45	42.45	40.86	44.86	37.72	44.46	37.60	44.18
		0.51	*0.71*	*0.70*	*0.70*	*0.71*	*0.93*	*0.84*	*0.94*	*0.79*	*1.05*
200	30	12.13	22.23	18.42	23.03	18.13	24.52	19.70	24.52	18.21	24.52
		0.25	*0.52*	*0.46*	*0.50*	*0.51*	*0.43*	*0.42*	*0.43*	*0.46*	*0.43*
200	35	18.92	39.74	39.51	39.55	38.69	33.69	33.15	34.14	35.45	32.49
		0.38	*0.53*	*0.53*	*0.54*	*0.52*	*0.67*	*0.63*	*0.67*	*0.55*	*0.66*
200	40	35.34	34.81	34.65	34.65	34.00	38.32	38.40	36.99	34.88	50.68
		0.73	*0.63*	*0.64*	*0.64*	*0.65*	*0.96*	*0.84*	*0.90*	*0.75*	*1.10*
200	45	29.58	35.33	35.27	35.33	53.13	35.33	34.50	35.31	36.98	35.33
		0.63	*0.73*	*0.77*	*0.73*	*1.25*	*0.73*	*0.79*	*0.73*	*0.85*	*0.73*
200	50	26.75	26.75	27.29	26.75	43.64	26.75	30.39	26.76	33.34	26.75
		0.53	*0.53*	*0.54*	*0.53*	*0.73*	*0.53*	*0.63*	*0.53*	*0.67*	*0.53*

Table C.4: Average Predictive Loss: Poisson, $p = 10, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.80	1.80	2.41	8.76	6.53	20.26	14.66	21.77	24.06	25.12
		0.17	*0.17*	*0.30*	*1.11*	*0.63*	*1.11*	*0.80*	*1.02*	*0.92*	*0.87*
100	2	1.50	7.86	5.84	5.72	4.05	5.48	4.17	5.44	5.23	5.47
		0.08	*0.16*	*0.22*	*0.17*	*0.20*	*0.17*	*0.18*	*0.16*	*0.17*	*0.17*
100	4	1.63	2.77	2.78	2.78	2.79	2.79	3.10	2.80	2.82	3.91
		0.08	*0.09*	*0.09*	*0.09*	*0.10*	*0.10*	*0.12*	*0.10*	*0.10*	*0.13*
100	6	2.80	3.82	3.88	3.92	3.73	4.04	3.81	3.97	3.74	4.31
		0.13	*0.19*	*0.19*	*0.19*	*0.18*	*0.19*	*0.17*	*0.19*	*0.18*	*0.18*
100	8	3.17	3.99	3.93	4.00	4.07	3.97	3.84	3.95	3.91	3.97
		0.13	*0.17*	*0.17*	*0.17*	*0.19*	*0.14*	*0.16*	*0.15*	*0.18*	*0.14*
100	10	16.15	18.13	18.97	17.26	21.62	16.15	18.49	16.78	20.66	16.15
		0.80	*0.80*	*0.82*	*0.78*	*0.81*	*0.80*	*0.79*	*0.81*	*0.82*	*0.80*
200	0	0.50	0.50	0.50	1.99	1.10	5.15	3.90	5.61	6.09	6.38
		0.04	*0.04*	*0.04*	*0.23*	*0.11*	*0.24*	*0.17*	*0.20*	*0.18*	*0.16*
200	2	1.65	1.72	1.72	1.72	2.03	1.75	3.12	1.72	2.03	4.12
		0.13	*0.13*	*0.13*	*0.13*	*0.14*	*0.13*	*0.16*	*0.13*	*0.14*	*0.16*
200	4	5.21	5.85	5.97	5.97	5.97	6.34	7.83	6.32	6.08	10.51
		0.25	*0.34*	*0.35*	*0.35*	*0.35*	*0.38*	*0.36*	*0.38*	*0.32*	*0.36*
200	6	3.67	6.28	6.13	6.13	6.10	5.95	5.14	5.92	5.89	5.41
		0.15	*0.25*	*0.25*	*0.25*	*0.24*	*0.24*	*0.21*	*0.23*	*0.24*	*0.19*
200	8	3.38	3.82	3.73	3.87	3.51	4.05	3.79	3.90	3.56	4.08
		0.12	*0.14*	*0.14*	*0.14*	*0.13*	*0.14*	*0.14*	*0.13*	*0.13*	*0.14*
200	10	7.15	7.15	7.22	7.15	7.67	7.15	7.23	14.10	7.46	7.15
		0.34	*0.34*	*0.34*	*0.34*	*0.36*	*0.34*	*0.34*	*0.40*	*0.35*	*0.34*
500	0	1.33	1.33	1.71	5.56	2.78	13.13	9.78	14.25	15.32	16.03
		0.11	*0.11*	*0.23*	*0.66*	*0.30*	*0.64*	*0.50*	*0.56*	*0.51*	*0.47*
500	2	1.36	1.41	1.41	1.41	1.56	1.42	2.79	1.41	1.57	3.98
		0.09	*0.10*	*0.10*	*0.10*	*0.11*	*0.11*	*0.14*	*0.10*	*0.11*	*0.14*
500	4	3.12	2.88	2.88	2.88	3.03	2.91	4.37	2.90	3.06	5.75
		0.19	*0.18*	*0.18*	*0.18*	*0.18*	*0.18*	*0.21*	*0.18*	*0.18*	*0.22*
500	6	5.15	2.93	2.92	2.92	3.38	2.96	5.92	2.95	3.32	8.13
		0.22	*0.17*	*0.17*	*0.17*	*0.20*	*0.19*	*0.28*	*0.19*	*0.20*	*0.29*
500	8	6.86	7.38	7.32	7.47	7.10	7.73	7.52	7.57	7.20	8.03
		0.30	*0.32*	*0.32*	*0.32*	*0.32*	*0.32*	*0.32*	*0.30*	*0.33*	*0.32*
500	10	3.36	3.48	3.62	3.44	4.76	3.37	3.52	5.61	4.34	3.36
		0.11	*0.12*	*0.12*	*0.12*	*0.17*	*0.11*	*0.12*	*0.12*	*0.15*	*0.11*

Table C.5: Average Predictive Loss: Poisson, $p = 20, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.13	1.13	1.32	7.40	5.78	27.03	17.05	26.77	31.25	31.87
		0.10	*0.10*	*0.18*	*1.13*	*0.47*	*1.20*	*0.76*	*1.16*	*0.91*	*0.87*
100	4	3.50	7.01	6.38	8.87	6.32	16.28	10.17	15.95	11.19	16.29
		0.16	*0.49*	*0.44*	*0.57*	*0.40*	*0.45*	*0.43*	*0.48*	*0.46*	*0.45*
100	8	1.95	2.45	2.50	2.50	2.53	2.91	3.73	2.79	2.73	5.09
		0.07	*0.11*	*0.11*	*0.11*	*0.11*	*0.15*	*0.13*	*0.14*	*0.11*	*0.14*
100	12	9.81	16.47	13.95	17.70	11.25	18.99	14.41	18.90	12.84	18.99
		0.45	*0.76*	*0.69*	*0.78*	*0.54*	*0.77*	*0.70*	*0.76*	*0.63*	*0.77*
100	16	6.55	8.81	9.43	8.62	11.54	8.28	8.93	8.32	9.98	8.28
		0.19	*0.25*	*0.26*	*0.25*	*0.30*	*0.22*	*0.24*	*0.23*	*0.26*	*0.22*
100	20	13.29	21.63	21.64	20.50	20.75	15.09	15.77	15.60	18.99	13.29
		0.38	*0.39*	*0.41*	*0.43*	*0.35*	*0.44*	*0.39*	*0.45*	*0.40*	*0.38*
200	0	2.73	2.73	3.54	9.73	11.05	52.00	36.73	56.83	66.93	68.05
		0.27	*0.27*	*0.46*	*1.78*	*1.01*	*2.86*	*1.68*	*2.59*	*2.02*	*1.97*
200	4	1.19	1.25	1.28	1.28	1.57	1.35	2.89	1.32	1.59	4.31
		0.07	*0.07*	*0.07*	*0.07*	*0.08*	*0.07*	*0.10*	*0.07*	*0.08*	*0.10*
200	8	3.16	3.59	3.67	3.67	3.78	4.03	5.44	3.97	4.14	7.53
		0.13	*0.16*	*0.16*	*0.16*	*0.16*	*0.18*	*0.20*	*0.17*	*0.17*	*0.22*
200	12	3.68	4.50	4.50	4.59	4.24	5.55	5.06	5.27	4.45	6.19
		0.11	*0.15*	*0.15*	*0.15*	*0.15*	*0.16*	*0.14*	*0.15*	*0.14*	*0.14*
200	16	5.52	7.14	7.08	7.19	7.22	6.95	7.07	7.02	7.19	6.89
		0.17	*0.18*	*0.18*	*0.19*	*0.17*	*0.19*	*0.18*	*0.19*	*0.18*	*0.19*
200	20	11.17	13.88	13.64	13.49	14.76	11.84	12.23	12.04	13.77	11.17
		0.29	*0.31*	*0.30*	*0.30*	*0.29*	*0.30*	*0.29*	*0.29*	*0.29*	*0.29*
500	0	0.73	0.73	0.94	3.75	2.97	16.45	12.06	17.17	20.21	20.55
		0.07	*0.07*	*0.13*	*0.60*	*0.26*	*0.70*	*0.44*	*0.64*	*0.42*	*0.39*
500	4	1.89	2.05	2.07	2.07	2.34	2.14	4.66	2.10	2.39	7.09
		0.09	*0.11*	*0.11*	*0.11*	*0.12*	*0.12*	*0.17*	*0.12*	*0.13*	*0.17*
500	8	3.76	4.10	4.14	4.14	4.19	4.31	6.24	4.31	4.39	8.28
		0.14	*0.16*	*0.16*	*0.16*	*0.16*	*0.17*	*0.19*	*0.17*	*0.17*	*0.20*
500	12	4.61	6.76	6.71	6.71	7.08	6.49	6.53	6.49	6.57	7.42
		0.15	*0.23*	*0.23*	*0.23*	*0.24*	*0.22*	*0.19*	*0.22*	*0.22*	*0.18*
500	16	11.38	15.06	14.77	14.78	16.22	14.32	13.66	14.22	14.35	14.13
		0.43	*0.57*	*0.55*	*0.55*	*0.59*	*0.51*	*0.45*	*0.50*	*0.53*	*0.45*
500	20	7.82	7.82	7.82	7.82	8.47	7.82	7.84	15.86	8.11	7.82
		0.20	*0.20*	*0.20*	*0.20*	*0.25*	*0.20*	*0.20*	*0.21*	*0.23*	*0.20*

Table C.6: Average Predictive Loss: Poisson, $p = 50, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
200	0	0.79	0.88	1.03	6.84	6.07	45.74	25.71	43.61	50.00	50.20
		0.07	*0.11*	*0.16*	*1.26*	*0.38*	*1.32*	*0.71*	*1.43*	*0.85*	*0.83*
200	5	2.94	3.25	3.35	3.35	5.85	3.62	16.97	3.56	6.90	38.11
		0.14	*0.16*	*0.17*	*0.17*	*0.30*	*0.18*	*0.64*	*0.18*	*0.37*	*0.84*
200	10	4.93	6.29	6.32	6.32	7.91	6.81	16.55	6.74	10.03	32.55
		0.18	*0.30*	*0.30*	*0.30*	*0.34*	*0.33*	*0.50*	*0.33*	*0.40*	*0.61*
200	15	10.71	11.04	11.36	11.36	13.25	12.16	27.09	12.10	17.46	52.62
		0.39	*0.41*	*0.43*	*0.43*	*0.48*	*0.45*	*0.80*	*0.45*	*0.64*	*1.18*
200	20	12.92	16.16	16.22	16.22	16.26	18.88	26.21	17.69	19.37	43.56
		0.34	*0.70*	*0.71*	*0.71*	*0.65*	*0.94*	*0.73*	*0.83*	*0.64*	*0.90*
200	25	18.57	27.03	26.59	26.59	25.90	40.54	30.83	36.92	25.89	47.56
		0.44	*0.94*	*0.90*	*0.90*	*0.90*	*1.32*	*0.88*	*1.36*	*0.79*	*1.03*
200	30	16.98	25.83	25.42	25.42	25.41	34.11	27.19	32.77	24.86	36.05
		0.35	*0.67*	*0.66*	*0.66*	*0.66*	*0.74*	*0.56*	*0.78*	*0.57*	*0.66*
200	35	20.48	30.95	30.91	30.91	29.63	31.53	29.77	31.00	29.24	34.89
		0.45	*0.44*	*0.42*	*0.42*	*0.45*	*0.57*	*0.61*	*0.55*	*0.52*	*0.69*
200	40	33.53	53.42	54.68	51.80	57.14	47.03	49.74	47.06	51.20	47.03
		0.89	*1.00*	*0.91*	*1.01*	*0.91*	*1.15*	*0.92*	*1.15*	*0.90*	*1.15*
200	45	46.29	56.48	62.25	55.66	71.95	54.77	56.68	54.76	60.44	54.77
		0.92	*1.15*	*1.20*	*1.13*	*1.10*	*1.08*	*1.04*	*1.08*	*1.07*	*1.08*
200	50	46.65	45.70	44.74	45.40	43.57	46.94	44.02	46.70	43.26	46.65
		0.99	*0.79*	*0.74*	*0.80*	*0.72*	*0.94*	*0.85*	*0.90*	*0.75*	*0.99*

Table C.7: Percent of Hits: Poisson, $p = 10, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.00	1.00	1.00	0.86	0.75	0.32	0.20	0.00	0.02	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*	*0.01*	*0.00*
100	2	1.00	0.95	0.95	0.95	0.77	0.92	0.30	0.92	0.76	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*
100	4	1.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
100	6	1.00	0.08	0.09	0.09	0.06	0.06	0.14	0.07	0.10	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.01*	*0.02*	*0.02*	*0.02*	*0.00*
100	8	1.00	0.03	0.07	0.02	0.00	0.00	0.10	0.00	0.04	0.00
		0.00	*0.01*	*0.02*	*0.01*	*0.00*	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*
100	10	1.00	0.37	0.10	0.57	0.01	0.94	0.08	0.00	0.02	1.00
		0.00	*0.03*	*0.02*	*0.03*	*0.01*	*0.02*	*0.02*	*0.00*	*0.01*	*0.00*
200	0	1.00	1.00	0.99	0.86	0.80	0.32	0.20	0.00	0.04	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*	*0.01*	*0.00*
200	2	1.00	0.97	0.97	0.97	0.88	0.94	0.26	0.96	0.85	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.01*	*0.03*	*0.01*	*0.02*	*0.00*
200	4	1.00	0.94	0.93	0.93	0.87	0.87	0.31	0.87	0.84	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
200	6	1.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
200	8	1.00	0.02	0.02	0.01	0.00	0.00	0.04	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
200	10	1.00	0.48	0.35	0.54	0.08	0.90	0.34	0.00	0.12	1.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.00*	*0.02*	*0.00*
500	0	1.00	1.00	0.98	0.81	0.85	0.33	0.17	0.00	0.04	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.02*	*0.03*	*0.02*	*0.00*	*0.01*	*0.00*
500	2	1.00	0.98	0.98	0.98	0.90	0.97	0.28	0.98	0.92	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.01*	*0.03*	*0.01*	*0.02*	*0.00*
500	4	1.00	0.97	0.96	0.96	0.93	0.93	0.34	0.93	0.91	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
500	6	1.00	0.90	0.90	0.90	0.96	0.84	0.56	0.84	0.90	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.01*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
500	8	1.00	0.82	0.83	0.80	0.94	0.66	0.72	0.61	0.91	0.00
		0.00	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
500	10	1.00	1.00	1.00	1.00	0.92	1.00	1.00	0.00	0.96	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*

Table C.8: Percent of Hits: Poisson, $p = 20, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.00	1.00	1.00	0.89	0.51	0.25	0.03	0.00	0.01	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.03*	*0.01*	*0.00*	*0.01*	*0.00*
100	4	1.00	0.93	0.91	0.91	0.63	0.84	0.08	0.85	0.50	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.02*	*0.03*	*0.00*
100	8	1.00	0.77	0.74	0.74	0.69	0.56	0.19	0.60	0.52	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
100	12	1.00	0.46	0.48	0.41	0.63	0.00	0.33	0.02	0.50	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.03*	*0.01*	*0.03*	*0.00*
100	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	20	1.00	0.80	0.03	0.90	0.00	1.00	0.00	0.00	0.00	1.00
		0.00	*0.03*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	0.99	0.91	0.70	0.33	0.04	0.00	0.00	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
200	4	1.00	0.92	0.92	0.92	0.76	0.89	0.09	0.89	0.67	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.02*	*0.03*	*0.00*
200	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	12	1.00	0.67	0.64	0.62	0.80	0.17	0.29	0.32	0.65	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*
200	16	1.00	0.02	0.04	0.01	0.00	0.00	0.06	0.00	0.03	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*
200	20	1.00	0.00	0.00	0.04	0.00	0.39	0.00	0.00	0.00	1.00
		0.00	*0.00*	*0.00*	*0.01*	*0.00*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	1.00	1.00	0.93	0.81	0.41	0.04	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
500	4	1.00	0.97	0.97	0.97	0.85	0.96	0.08	0.96	0.80	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.01*	*0.02*	*0.01*	*0.03*	*0.00*
500	8	1.00	0.76	0.77	0.77	0.76	0.72	0.11	0.72	0.70	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
500	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	16	1.00	0.64	0.64	0.60	0.94	0.09	0.49	0.16	0.88	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
500	20	1.00	1.00	1.00	1.00	0.78	1.00	0.98	0.00	0.93	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*	*0.01*	*0.00*	*0.02*	*0.00*

Table C.9: Percent of Hits: Poisson, $p = 50, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
200	0	1.00	1.00	1.00	0.92	0.40	0.14	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
200	5	1.00	0.92	0.92	0.92	0.38	0.84	0.00	0.85	0.30	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.00*	*0.02*	*0.03*	*0.00*
200	10	1.00	0.74	0.73	0.73	0.44	0.60	0.00	0.61	0.22	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.03*	*0.03*	*0.00*
200	15	1.00	0.81	0.79	0.79	0.70	0.62	0.05	0.67	0.39	0.00
		0.00	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.01*	*0.03*	*0.03*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.12	0.23	0.07	0.32	0.00	0.04	0.00	0.23	0.00
		0.00	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*	*0.01*	*0.00*	*0.03*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	50	1.00	1.00	0.02	1.00	0.00	1.00	0.00	0.00	0.00	1.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Table C.10: Percent of Hits: Poisson, $p = 10, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.00	1.00	0.98	0.84	0.71	0.32	0.18	0.00	0.02	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.02*	*0.00*	*0.01*	*0.00*
100	2	1.00	0.12	0.29	0.08	0.44	0.00	0.26	0.00	0.05	0.00
		0.00	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*	*0.03*	*0.00*	*0.01*	*0.00*
100	4	1.00	0.03	0.03	0.03	0.10	0.07	0.14	0.06	0.11	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.00*
100	6	1.00	0.69	0.67	0.66	0.73	0.53	0.48	0.54	0.72	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
100	8	1.00	0.53	0.63	0.40	0.73	0.00	0.64	0.02	0.74	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.03*	*0.01*	*0.03*	*0.00*
100	10	1.00	0.58	0.07	0.74	0.00	1.00	0.02	0.00	0.00	1.00
		0.00	*0.03*	*0.02*	*0.03*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	1.00	0.86	0.86	0.32	0.17	0.00	0.03	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.02*	*0.03*	*0.02*	*0.00*	*0.01*	*0.00*
200	2	1.00	0.97	0.97	0.97	0.81	0.96	0.21	0.97	0.81	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.01*	*0.03*	*0.01*	*0.02*	*0.00*
200	4	1.00	0.92	0.91	0.91	0.90	0.87	0.43	0.87	0.86	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
200	6	1.00	0.42	0.44	0.44	0.43	0.42	0.46	0.42	0.47	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
200	8	1.00	0.70	0.76	0.64	0.96	0.26	0.67	0.34	0.92	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.01*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
200	10	1.00	1.00	0.98	1.00	0.91	1.00	0.98	0.04	0.94	1.00
		0.00	*0.00*	*0.01*	*0.00*	*0.02*	*0.00*	*0.01*	*0.01*	*0.02*	*0.00*
500	0	1.00	1.00	0.98	0.84	0.88	0.32	0.22	0.00	0.04	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.02*	*0.03*	*0.03*	*0.00*	*0.01*	*0.00*
500	2	1.00	0.98	0.98	0.98	0.92	0.98	0.27	0.98	0.92	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.01*	*0.03*	*0.01*	*0.02*	*0.00*
500	4	1.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
500	6	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	8	1.00	0.84	0.85	0.80	0.97	0.64	0.72	0.60	0.93	0.00
		0.00	*0.02*	*0.02*	*0.03*	*0.01*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
500	10	1.00	0.94	0.86	0.95	0.49	0.98	0.89	0.00	0.58	1.00
		0.00	*0.02*	*0.02*	*0.01*	*0.03*	*0.01*	*0.02*	*0.00*	*0.03*	*0.00*

Table C.11: Percent of Hits: Poisson, $p = 20, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
100	0	1.00	1.00	0.99	0.88	0.58	0.25	0.05	0.00	0.00	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
100	4	1.00	0.70	0.72	0.61	0.64	0.00	0.09	0.05	0.09	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.02*	*0.01*	*0.02*	*0.00*
100	8	1.00	0.80	0.78	0.78	0.72	0.68	0.13	0.69	0.62	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
100	12	1.00	0.46	0.63	0.30	0.85	0.00	0.44	0.00	0.69	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*	*0.03*	*0.00*	*0.03*	*0.00*
100	16	1.00	0.00	0.11	0.00	0.05	0.00	0.16	0.00	0.10	0.00
		0.00	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*	*0.02*	*0.00*	*0.02*	*0.00*
100	20	1.00	0.16	0.00	0.26	0.00	0.84	0.00	0.00	0.00	1.00
		0.00	*0.02*	*0.00*	*0.03*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	0.98	0.93	0.69	0.36	0.03	0.00	0.00	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
200	4	1.00	0.98	0.96	0.96	0.75	0.92	0.07	0.93	0.72	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.02*	*0.02*	*0.02*	*0.03*	*0.00*
200	8	1.00	0.87	0.85	0.85	0.78	0.72	0.17	0.74	0.63	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
200	12	1.00	0.67	0.66	0.65	0.81	0.41	0.28	0.46	0.67	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
200	16	1.00	0.01	0.02	0.02	0.00	0.00	0.04	0.00	0.01	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*
200	20	1.00	0.08	0.00	0.12	0.00	0.81	0.00	0.00	0.00	1.00
		0.00	*0.02*	*0.00*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	1.00	0.98	0.90	0.73	0.29	0.07	0.00	0.00	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*
500	4	1.00	0.95	0.95	0.95	0.85	0.93	0.09	0.94	0.84	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.00*
500	8	1.00	0.90	0.89	0.89	0.86	0.83	0.14	0.83	0.79	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.00*
500	12	1.00	0.48	0.47	0.47	0.45	0.45	0.21	0.44	0.49	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
500	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	20	1.00	1.00	1.00	1.00	0.90	1.00	1.00	0.00	0.95	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*

Table C.12: Percent of Hits: Poisson, $p = 50, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	hcml	aic	cml	mcml	ns
200	0	1.00	1.00	1.00	0.92	0.40	0.14	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
200	5	1.00	0.92	0.92	0.92	0.38	0.84	0.00	0.85	0.30	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.00*	*0.02*	*0.03*	*0.00*
200	10	1.00	0.74	0.73	0.73	0.44	0.60	0.00	0.61	0.22	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.03*	*0.03*	*0.00*
200	15	1.00	0.81	0.79	0.79	0.70	0.62	0.05	0.67	0.39	0.00
		0.00	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.01*	*0.03*	*0.03*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.12	0.23	0.07	0.32	0.00	0.04	0.00	0.23	0.00
		0.00	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*	*0.01*	*0.00*	*0.03*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	50	1.00	1.00	0.02	1.00	0.00	1.00	0.00	0.00	0.00	1.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

C.2 Logistic Models

Table C.13: Average Predictive Loss: Logistic, $p = 10, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	0.26	0.40	0.39	0.77	1.73	1.81	2.49	2.69	2.76	2.77
		0.02	*0.06*	*0.06*	*0.06*	*0.08*	*0.12*	*0.09*	*0.08*	*0.07*	*0.07*
100	2	0.59	2.24	2.78	1.57	1.86	2.36	2.33	2.34	2.34	2.34
		0.03	*0.06*	*0.04*	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	4	0.83	1.93	2.12	1.64	1.81	2.08	2.07	2.08	2.08	2.08
		0.04	*0.06*	*0.06*	*0.05*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	6	1.35	2.31	2.25	3.10	2.19	2.17	2.16	2.17	2.17	2.17
		0.05	*0.09*	*0.08*	*0.10*	*0.08*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	8	1.73	2.23	2.20	2.11	2.18	2.16	2.16	2.16	2.16	2.16
		0.05	*0.06*	*0.07*	*0.05*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	10	2.16	2.36	2.16	4.20	2.72	2.16	2.19	2.16	2.16	2.16
		0.06	*0.07*	*0.06*	*0.09*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	0	0.20	0.24	0.23	0.51	1.40	1.21	2.06	2.23	2.29	2.29
		0.02	*0.03*	*0.03*	*0.05*	*0.06*	*0.10*	*0.07*	*0.06*	*0.06*	*0.06*
200	2	0.62	2.50	3.28	1.69	1.69	2.42	2.36	2.34	2.37	2.37
		0.03	*0.09*	*0.07*	*0.08*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	4	0.99	1.94	2.19	1.91	1.83	2.13	2.11	1.99	2.12	2.12
		0.04	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	6	1.29	2.05	2.06	2.28	1.96	2.02	2.02	2.00	2.02	2.02
		0.05	*0.06*	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	8	1.66	1.93	2.06	1.74	1.88	2.07	2.06	1.99	2.07	2.07
		0.05	*0.06*	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	10	2.08	2.21	2.08	3.81	2.47	2.08	2.16	2.16	2.08	2.08
		0.06	*0.06*	*0.06*	*0.08*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
500	0	0.16	0.17	0.16	0.30	1.14	0.81	1.72	1.82	1.87	1.87
		0.01	*0.02*	*0.01*	*0.03*	*0.05*	*0.08*	*0.06*	*0.05*	*0.05*	*0.05*
500	2	0.77	0.48	0.44	0.66	1.79	1.41	2.53	2.69	2.74	2.74
		0.04	*0.04*	*0.04*	*0.05*	*0.08*	*0.11*	*0.09*	*0.08*	*0.08*	*0.08*
500	4	0.92	1.88	2.12	2.16	1.73	2.17	2.19	1.76	2.16	2.15
		0.04	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.06*	*0.07*	*0.06*	*0.06*
500	6	1.21	1.90	2.00	2.10	1.82	1.96	1.93	1.84	1.95	1.95
		0.04	*0.06*	*0.05*	*0.05*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
500	8	1.58	2.02	1.99	2.73	2.01	1.99	1.97	2.01	1.99	1.99
		0.04	*0.06*	*0.05*	*0.08*	*0.06*	*0.05*	*0.05*	*0.06*	*0.05*	*0.05*
500	10	2.10	2.40	2.13	4.46	2.58	2.10	2.15	2.46	2.10	2.10
		0.06	*0.07*	*0.06*	*0.10*	*0.07*	*0.06*	*0.06*	*0.07*	*0.06*	*0.06*

Table C.14: Average Predictive Loss: Logistic, $p = 20, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	0.20	0.25	0.71	1.09	2.94	4.23	4.64	4.55	4.71	4.71
		0.01	*0.04*	*0.12*	*0.08*	*0.10*	*0.14*	*0.10*	*0.10*	*0.09*	*0.09*
100	4	1.20	1.38	2.44	2.21	3.67	4.90	4.97	4.95	4.99	5.00
		0.05	*0.06*	*0.15*	*0.07*	*0.10*	*0.10*	*0.09*	*0.10*	*0.09*	*0.09*
100	8	1.70	4.76	4.41	4.57	3.97	4.41	4.40	4.40	4.41	4.41
		0.05	*0.12*	*0.09*	*0.10*	*0.11*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*
100	12	2.57	4.71	4.59	4.41	4.16	4.59	4.58	4.55	4.59	4.59
		0.07	*0.11*	*0.11*	*0.10*	*0.11*	*0.11*	*0.10*	*0.10*	*0.11*	*0.11*
100	16	3.58	4.82	4.51	5.20	4.63	4.51	4.51	4.48	4.51	4.51
		0.09	*0.10*	*0.10*	*0.08*	*0.09*	*0.10*	*0.10*	*0.09*	*0.10*	*0.10*
100	20	4.39	4.82	4.39	6.10	4.82	4.39	4.39	4.39	4.39	4.39
		0.10	*0.13*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*	*0.10*
200	0	0.25	0.29	0.33	1.09	3.26	3.96	4.82	5.22	5.32	5.32
		0.02	*0.04*	*0.06*	*0.08*	*0.11*	*0.20*	*0.14*	*0.11*	*0.10*	*0.10*
200	4	1.08	4.72	5.44	3.66	3.90	4.99	4.98	4.98	4.99	4.99
		0.05	*0.09*	*0.07*	*0.11*	*0.11*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*
200	8	1.65	3.83	4.18	3.49	3.42	4.18	4.18	4.18	4.18	4.18
		0.05	*0.10*	*0.09*	*0.12*	*0.10*	*0.09*	*0.08*	*0.09*	*0.09*	*0.09*
200	12	2.49	4.25	4.13	5.38	4.26	4.13	4.13	4.13	4.13	4.13
		0.06	*0.08*	*0.08*	*0.09*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
200	16	3.38	4.39	4.29	8.62	5.12	4.29	4.31	4.29	4.29	4.29
		0.07	*0.08*	*0.08*	*0.14*	*0.10*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
200	20	4.00	4.09	4.00	7.44	4.93	4.00	4.02	4.00	4.00	4.00
		0.07	*0.07*	*0.07*	*0.11*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
500	0	0.21	0.27	0.26	0.73	3.07	2.50	4.21	5.08	5.15	5.15
		0.02	*0.03*	*0.05*	*0.07*	*0.10*	*0.20*	*0.16*	*0.11*	*0.10*	*0.10*
500	4	0.95	1.58	2.26	1.30	2.95	3.36	4.33	3.41	4.34	4.34
		0.04	*0.07*	*0.13*	*0.06*	*0.09*	*0.14*	*0.08*	*0.09*	*0.08*	*0.08*
500	8	1.62	1.97	2.39	1.67	3.02	3.00	3.96	3.25	3.99	3.99
		0.05	*0.07*	*0.11*	*0.06*	*0.08*	*0.12*	*0.08*	*0.09*	*0.08*	*0.08*
500	12	2.50	3.95	4.08	5.19	3.80	4.08	4.07	3.89	4.08	4.08
		0.06	*0.08*	*0.08*	*0.13*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
500	16	3.12	4.10	3.94	5.32	4.17	3.94	3.93	4.03	3.94	3.94
		0.06	*0.07*	*0.07*	*0.10*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
500	20	3.99	4.04	3.99	6.81	4.41	3.99	4.10	4.15	3.99	3.99
		0.09	*0.09*	*0.09*	*0.18*	*0.10*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*

Table C.15: Average Predictive Loss: Logistic, $p = 50, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
200	0	0.24	0.31	7.45	2.11	8.23	13.00	12.99	12.75	13.00	13.00
		0.02	*0.05*	*0.47*	*0.12*	*0.18*	*0.15*	*0.15*	*0.16*	*0.15*	*0.15*
200	5	1.26	7.23	12.19	5.16	8.57	12.19	12.17	12.16	12.17	12.19
		0.04	*0.20*	*0.14*	*0.14*	*0.16*	*0.14*	*0.14*	*0.14*	*0.14*	*0.14*
200	10	2.05	10.55	11.28	6.52	8.49	11.28	11.26	11.11	11.27	11.28
		0.05	*0.21*	*0.15*	*0.15*	*0.17*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	15	2.96	11.01	11.02	8.59	9.08	11.02	11.00	10.87	11.01	11.02
		0.06	*0.19*	*0.17*	*0.14*	*0.18*	*0.17*	*0.17*	*0.16*	*0.17*	*0.17*
200	20	4.03	8.28	11.07	4.57	8.17	11.07	11.05	10.92	11.05	11.07
		0.08	*0.29*	*0.15*	*0.13*	*0.16*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	25	5.23	10.05	11.18	5.73	8.65	11.18	11.16	10.93	11.17	11.18
		0.09	*0.22*	*0.15*	*0.11*	*0.15*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	30	6.21	11.56	11.61	8.66	9.64	11.61	11.59	11.22	11.60	11.61
		0.10	*0.21*	*0.22*	*0.17*	*0.23*	*0.22*	*0.22*	*0.21*	*0.22*	*0.22*
200	35	7.13	11.41	10.95	12.49	10.40	10.95	10.93	10.88	10.94	10.95
		0.11	*0.16*	*0.14*	*0.12*	*0.13*	*0.14*	*0.14*	*0.14*	*0.14*	*0.14*
200	40	8.47	11.60	11.48	13.09	11.39	11.48	11.48	11.33	11.49	11.49
		0.12	*0.20*	*0.19*	*0.14*	*0.19*	*0.19*	*0.19*	*0.19*	*0.19*	*0.19*
200	45	9.68	11.22	11.05	13.05	11.07	11.05	10.97	10.91	11.00	11.06
		0.14	*0.17*	*0.16*	*0.13*	*0.15*	*0.16*	*0.16*	*0.15*	*0.16*	*0.16*
200	50	11.32	11.47	11.32	11.94	10.85	11.32	11.32	11.22	11.31	11.32
		0.16	*0.17*	*0.16*	*0.12*	*0.14*	*0.16*	*0.16*	*0.15*	*0.16*	*0.16*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table C.16: Average Predictive Loss: Logistic, $p = 10, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	0.26	0.41	0.40	0.76	1.80	2.02	2.61	2.75	2.83	2.83
		0.02	*0.06*	*0.07*	*0.07*	*0.09*	*0.12*	*0.09*	*0.08*	*0.08*	*0.08*
100	2	0.73	0.39	0.41	0.82	1.75	1.89	2.54	2.66	2.75	2.75
		0.04	*0.06*	*0.07*	*0.07*	*0.09*	*0.12*	*0.09*	*0.08*	*0.08*	*0.08*
100	4	0.93	2.28	2.16	2.39	2.01	2.12	2.12	2.12	2.12	2.12
		0.04	*0.07*	*0.06*	*0.08*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	6	1.27	2.29	2.07	2.76	2.22	2.06	2.07	2.06	2.06	2.06
		0.04	*0.07*	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	8	1.75	2.34	2.17	2.68	2.29	2.17	2.17	2.17	2.17	2.17
		0.05	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
100	10	2.21	2.11	2.19	1.88	2.13	2.21	2.21	2.21	2.21	2.21
		0.06	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	0	0.26	0.33	0.26	0.57	1.66	1.38	2.45	2.65	2.75	2.75
		0.02	*0.04*	*0.02*	*0.06*	*0.08*	*0.12*	*0.09*	*0.08*	*0.07*	*0.07*
200	2	0.70	1.91	1.95	1.77	2.12	2.58	2.61	2.65	2.65	2.65
		0.04	*0.04*	*0.05*	*0.04*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
200	4	0.94	1.45	1.69	1.25	1.69	1.98	2.10	1.92	2.12	2.12
		0.04	*0.06*	*0.07*	*0.05*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	6	1.34	1.79	1.94	1.55	1.95	2.13	2.16	2.07	2.16	2.16
		0.05	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	8	1.65	2.07	2.04	2.61	1.97	2.03	2.02	2.03	2.03	2.03
		0.05	*0.06*	*0.06*	*0.10*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
200	10	2.10	2.33	2.10	4.47	2.69	2.10	2.18	2.24	2.10	2.10
		0.06	*0.06*	*0.06*	*0.09*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*	*0.06*
500	0	0.20	0.23	0.23	0.39	1.52	1.08	2.19	2.37	2.43	2.43
		0.02	*0.03*	*0.03*	*0.04*	*0.07*	*0.10*	*0.08*	*0.07*	*0.07*	*0.07*
500	2	0.71	2.42	3.38	1.98	2.04	2.61	2.63	2.47	2.64	2.64
		0.03	*0.13*	*0.13*	*0.10*	*0.08*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*
500	4	0.87	1.38	1.54	1.10	1.56	1.76	2.00	1.56	2.02	2.02
		0.04	*0.07*	*0.08*	*0.06*	*0.06*	*0.07*	*0.06*	*0.06*	*0.06*	*0.06*
500	6	1.27	2.39	2.39	2.63	2.13	2.22	2.10	2.13	2.08	2.07
		0.05	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*	*0.06*	*0.07*	*0.06*	*0.06*
500	8	1.62	2.13	2.04	3.36	2.20	2.04	2.06	2.17	2.04	2.04
		0.05	*0.06*	*0.05*	*0.10*	*0.07*	*0.05*	*0.06*	*0.06*	*0.05*	*0.05*
500	10	2.05	2.43	2.20	3.42	2.43	2.06	2.05	2.35	2.05	2.05
		0.06	*0.06*	*0.06*	*0.08*	*0.06*	*0.06*	*0.06*	*0.05*	*0.06*	*0.06*

Table C.17: Average Predictive Loss: Logistic, $p = 20, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	0.31	0.44	0.86	1.31	3.31	4.93	5.27	5.15	5.34	5.34
		0.03	*0.06*	*0.14*	*0.09*	*0.11*	*0.15*	*0.10*	*0.11*	*0.10*	*0.10*
100	4	1.12	2.89	3.87	2.85	3.75	4.95	4.94	4.92	4.95	4.96
		0.04	*0.06*	*0.11*	*0.07*	*0.09*	*0.08*	*0.08*	*0.09*	*0.08*	*0.08*
100	8	1.79	4.57	4.46	4.26	3.90	4.46	4.44	4.44	4.45	4.46
		0.05	*0.12*	*0.11*	*0.11*	*0.11*	*0.11*	*0.11*	*0.10*	*0.11*	*0.11*
100	12	2.58	5.26	4.47	4.74	4.19	4.47	4.46	4.46	4.47	4.47
		0.06	*0.12*	*0.09*	*0.10*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*	*0.09*
100	16	3.62	4.65	4.56	4.12	4.23	4.56	4.54	4.52	4.55	4.56
		0.08	*0.11*	*0.10*	*0.09*	*0.10*	*0.10*	*0.10*	*0.09*	*0.10*	*0.10*
100	20	4.35	5.19	4.39	6.63	5.26	4.35	4.37	4.35	4.34	4.35
		0.08	*0.11*	*0.09*	*0.06*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
200	0	0.24	0.35	0.33	1.04	3.31	4.16	5.00	5.25	5.37	5.37
		0.02	*0.05*	*0.06*	*0.08*	*0.11*	*0.19*	*0.13*	*0.11*	*0.10*	*0.10*
200	4	1.11	2.32	2.58	2.65	3.60	4.78	4.85	4.88	4.90	4.90
		0.04	*0.05*	*0.10*	*0.06*	*0.09*	*0.10*	*0.09*	*0.09*	*0.09*	*0.09*
200	8	1.55	3.87	4.04	4.05	3.53	4.04	4.04	4.04	4.04	4.04
		0.04	*0.09*	*0.07*	*0.11*	*0.09*	*0.07*	*0.07*	*0.07*	*0.07*	*0.07*
200	12	2.36	2.29	3.99	1.58	2.86	4.08	4.08	4.01	4.08	4.08
		0.06	*0.11*	*0.09*	*0.08*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
200	16	3.36	4.76	4.28	5.86	4.68	4.28	4.28	4.28	4.28	4.28
		0.08	*0.09*	*0.08*	*0.09*	*0.09*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
200	20	4.07	4.26	4.07	6.59	4.67	4.07	4.08	4.08	4.07	4.07
		0.08	*0.09*	*0.08*	*0.10*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
500	0	0.25	0.27	0.25	0.68	2.90	2.21	4.09	4.89	4.99	4.99
		0.02	*0.03*	*0.02*	*0.06*	*0.10*	*0.19*	*0.15*	*0.10*	*0.09*	*0.09*
500	4	1.12	4.09	5.37	3.76	3.55	4.75	4.75	4.57	4.75	4.75
		0.04	*0.15*	*0.15*	*0.12*	*0.10*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*
500	8	1.72	3.28	4.08	3.56	3.22	4.06	4.05	3.52	4.06	4.06
		0.05	*0.11*	*0.08*	*0.13*	*0.09*	*0.08*	*0.08*	*0.09*	*0.08*	*0.08*
500	12	2.29	3.88	3.93	5.04	3.71	3.93	3.92	3.80	3.93	3.93
		0.06	*0.09*	*0.08*	*0.13*	*0.09*	*0.08*	*0.08*	*0.09*	*0.08*	*0.08*
500	16	3.27	4.09	4.05	5.01	4.15	4.05	4.06	4.10	4.05	4.05
		0.07	*0.08*	*0.08*	*0.12*	*0.09*	*0.08*	*0.08*	*0.09*	*0.08*	*0.08*
500	20	4.01	3.82	4.01	3.31	3.49	4.01	4.00	3.73	4.01	4.01
		0.08	*0.08*	*0.08*	*0.12*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*	*0.08*

Table C.18: Average Predictive Loss: Logistic, $p = 50, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
200	0	0.23	0.26	7.61	1.79	7.73	12.79	12.78	12.49	12.78	12.79
		0.02	*0.03*	*0.46*	*0.11*	*0.18*	*0.14*	*0.14*	*0.15*	*0.14*	*0.14*
200	5	1.49	0.90	9.00	2.82	8.34	13.04	13.03	12.80	13.03	13.04
		0.05	*0.04*	*0.43*	*0.13*	*0.18*	*0.15*	*0.15*	*0.16*	*0.15*	*0.15*
200	10	2.15	6.92	11.35	3.32	7.66	11.35	11.33	11.17	11.34	11.35
		0.05	*0.35*	*0.15*	*0.14*	*0.17*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	15	3.17	8.87	11.63	4.81	8.51	11.63	11.61	11.42	11.62	11.63
		0.07	*0.31*	*0.17*	*0.12*	*0.18*	*0.17*	*0.17*	*0.16*	*0.17*	*0.17*
200	20	4.10	10.93	11.38	6.94	9.12	11.38	11.37	11.18	11.38	11.38
		0.08	*0.21*	*0.17*	*0.12*	*0.18*	*0.17*	*0.17*	*0.16*	*0.17*	*0.17*
200	25	4.99	9.43	10.91	5.97	8.48	10.91	10.89	10.73	10.90	10.91
		0.08	*0.23*	*0.15*	*0.12*	*0.16*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	30	6.24	10.36	11.39	6.41	9.23	11.39	11.37	11.13	11.38	11.39
		0.10	*0.25*	*0.18*	*0.12*	*0.18*	*0.18*	*0.18*	*0.18*	*0.19*	*0.18*
200	35	7.43	11.47	11.38	8.53	9.77	11.38	11.37	11.21	11.37	11.38
		0.12	*0.21*	*0.19*	*0.10*	*0.19*	*0.19*	*0.19*	*0.19*	*0.19*	*0.19*
200	40	8.64	11.45	11.25	9.74	9.97	11.25	11.23	11.12	11.24	11.25
		0.12	*0.15*	*0.15*	*0.11*	*0.14*	*0.15*	*0.15*	*0.14*	*0.15*	*0.15*
200	45	9.99	11.62	11.46	9.81	10.16	11.46	11.39	11.31	11.42	11.46
		0.14	*0.19*	*0.19*	*0.11*	*0.18*	*0.19*	*0.19*	*0.18*	*0.19*	*0.19*
200	50	11.60	11.53	11.59	9.65	10.39	11.59	11.57	11.44	11.58	11.60
		0.15	*0.16*	*0.15*	*0.10*	*0.14*	*0.15*	*0.15*	*0.15*	*0.15*	*0.15*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table C.19: Percent of Hits: Logistic, $p = 10, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	1.00	0.97	0.98	0.72	0.16	0.55	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.03*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
100	2	1.00	0.10	0.01	0.24	0.14	0.00	0.00	0.00	0.00	0.00
		0.00	*0.02*	*0.01*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	4	1.00	0.03	0.00	0.07	0.07	0.00	0.00	0.01	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
100	6	1.00	0.10	0.00	0.14	0.27	0.00	0.00	0.00	0.00	0.00
		0.00	*0.02*	*0.00*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	8	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	10	1.00	0.20	1.00	0.00	0.02	1.00	0.00	1.00	1.00	1.00
		0.00	*0.03*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	0	1.00	0.99	1.00	0.80	0.16	0.68	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.03*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
200	2	1.00	0.28	0.12	0.44	0.22	0.00	0.00	0.01	0.00	0.00
		0.00	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	4	1.00	0.26	0.03	0.32	0.22	0.00	0.00	0.09	0.00	0.00
		0.00	*0.03*	*0.01*	*0.03*	*0.03*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*
200	6	1.00	0.06	0.00	0.01	0.09	0.00	0.00	0.07	0.00	0.00
		0.00	*0.02*	*0.00*	*0.01*	*0.02*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*
200	8	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	10	1.00	0.25	1.00	0.00	0.03	1.00	0.00	0.28	1.00	1.00
		0.00	*0.03*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*
500	0	1.00	0.99	1.00	0.90	0.20	0.77	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
500	2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	4	1.00	0.34	0.29	0.32	0.29	0.20	0.01	0.28	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*
500	6	1.00	0.24	0.01	0.10	0.32	0.00	0.00	0.29	0.00	0.00
		0.00	*0.03*	*0.01*	*0.02*	*0.03*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*
500	8	1.00	0.36	0.00	0.25	0.49	0.00	0.00	0.44	0.00	0.00
		0.00	*0.03*	*0.00*	*0.03*	*0.03*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*
500	10	1.00	0.16	0.98	0.00	0.02	1.00	0.00	0.06	1.00	1.00
		0.00	*0.02*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*

Table C.20: Percent of Hits: Logistic, $p = 20, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	1.00	0.99	0.93	0.53	0.03	0.19	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.02*	*0.03*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
100	4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	8	1.00	0.01	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	20	1.00	0.06	1.00	0.00	0.00	1.00	0.00	0.70	0.81	0.99
		0.00	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.03*	*0.02*	*0.01*
200	0	1.00	0.99	0.99	0.60	0.02	0.38	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
200	4	1.00	0.04	0.00	0.09	0.02	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	8	1.00	0.09	0.00	0.26	0.12	0.00	0.00	0.00	0.00	0.00
		0.00	*0.02*	*0.00*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	12	1.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	16	1.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.06	1.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00
		0.00	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	0.98	1.00	0.77	0.02	0.65	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
500	4	1.00	0.69	0.66	0.82	0.03	0.39	0.00	0.01	0.00	0.00
		0.00	*0.03*	*0.03*	*0.02*	*0.01*	*0.03*	*0.00*	*0.01*	*0.00*	*0.00*
500	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	12	1.00	0.05	0.00	0.03	0.12	0.00	0.00	0.05	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
500	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	20	1.00	0.38	1.00	0.00	0.03	1.00	0.00	0.19	1.00	1.00
		0.00	*0.03*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*

Table C.21: Percent of Hits: Logistic, $p = 50, \rho = 0$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
200	0	1.00	0.99	0.51	0.32	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	5	1.00	0.05	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	10	1.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	15	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.06	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*
200	50	1.00	0.02	1.00	0.00	0.00	1.00	0.00	0.01	0.43	1.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.03*	*0.00*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

Table C.22: Percent of Hits: Logistic, $p = 10, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	1.00	0.96	0.98	0.74	0.20	0.48	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.03*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
100	2	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	4	1.00	0.10	0.00	0.12	0.16	0.00	0.00	0.00	0.00	0.00
		0.00	*0.02*	*0.00*	*0.02*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	6	1.00	0.04	0.00	0.01	0.03	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	10	1.00	0.04	0.95	0.00	0.00	1.00	0.00	0.94	1.00	1.00
		0.00	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	0	1.00	0.98	1.00	0.85	0.20	0.72	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
200	2	1.00	0.02	0.00	0.05	0.08	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	4	1.00	0.01	0.00	0.01	0.04	0.00	0.00	0.02	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	6	1.00	0.01	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	8	1.00	0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	10	1.00	0.18	1.00	0.00	0.01	1.00	0.00	0.20	1.00	1.00
		0.00	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*
500	0	1.00	0.99	1.00	0.90	0.16	0.76	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
500	2	1.00	0.44	0.27	0.50	0.24	0.11	0.02	0.08	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.01*	*0.02*	*0.00*	*0.00*
500	4	1.00	0.65	0.62	0.90	0.33	0.44	0.06	0.36	0.00	0.00
		0.00	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*
500	6	1.00	0.01	0.00	0.00	0.04	0.00	0.00	0.03	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
500	8	1.00	0.26	0.00	0.04	0.30	0.00	0.00	0.29	0.00	0.00
		0.00	*0.03*	*0.00*	*0.01*	*0.03*	*0.00*	*0.00*	*0.03*	*0.00*	*0.00*
500	10	1.00	0.05	0.87	0.00	0.00	0.98	0.00	0.01	1.00	1.00
		0.00	*0.01*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*

Table C.23: Percent of Hits: Logistic, $p = 20, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
100	0	1.00	0.97	0.93	0.57	0.03	0.14	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.02*	*0.03*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*
100	4	1.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	8	1.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.01*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	20	1.00	0.10	0.99	0.00	0.00	1.00	0.00	0.90	0.92	1.00
		0.00	*0.02*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.02*	*0.00*
200	0	1.00	0.98	0.99	0.64	0.03	0.34	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
200	4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	8	1.00	0.04	0.00	0.06	0.08	0.00	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.02*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.01	1.00	0.00	0.00	1.00	0.00	1.00	1.00	1.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	0.99	1.00	0.80	0.05	0.70	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.00*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
500	4	1.00	0.18	0.06	0.18	0.05	0.00	0.00	0.00	0.00	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	8	1.00	0.27	0.01	0.32	0.12	0.00	0.00	0.05	0.00	0.00
		0.00	*0.03*	*0.01*	*0.03*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
500	12	1.00	0.03	0.00	0.03	0.12	0.00	0.00	0.02	0.00	0.00
		0.00	*0.01*	*0.00*	*0.01*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
500	16	1.00	0.02	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00
		0.00	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
500	20	1.00	0.01	1.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Table C.24: Percent of Hits: Logistic, $p = 50, \rho = 0.5$

n	q	gold	fb.r	cfb	bic	aic	fb	cml	mcml	hcml	ns
200	0	1.00	1.00	0.48	0.41	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	5	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	10	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	15	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.04	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*
200	50	1.00	0.01	1.00	0.00	0.00	1.00	0.00	0.04	0.53	1.00
		0.00	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.03*	*0.00*

Note: For $p = 50, n = 200$, there were many warnings of convergence problem.

C.3 Normal Linear Models

Table C.25: Average Predictive Loss: Normal, $p = 10, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.06	1.06	1.26	3.61	2.94	6.61	9.17	10.19	8.79	10.83
		0.09	*0.09*	*0.13*	*0.40*	*0.22*	*0.28*	*0.36*	*0.31*	*0.39*	*0.26*
100	2	2.96	3.87	4.07	4.49	4.36	7.41	6.47	7.03	8.41	10.75
		0.15	*0.24*	*0.24*	*0.30*	*0.24*	*0.30*	*0.38*	*0.31*	*0.39*	*0.29*
100	4	5.08	11.41	10.43	11.40	10.16	9.88	11.63	9.89	11.46	11.43
		0.19	*0.35*	*0.35*	*0.32*	*0.34*	*0.34*	*0.28*	*0.34*	*0.28*	*0.28*
100	6	7.24	10.44	9.85	11.25	8.85	10.18	11.48	10.25	11.55	11.55
		0.25	*0.32*	*0.31*	*0.31*	*0.28*	*0.31*	*0.31*	*0.31*	*0.31*	*0.31*
100	8	9.58	11.51	12.62	11.51	15.73	12.46	11.49	12.11	11.51	11.51
		0.27	*0.30*	*0.40*	*0.30*	*0.53*	*0.36*	*0.30*	*0.34*	*0.30*	*0.30*
100	10	11.37	12.72	13.15	11.87	13.98	12.36	11.45	12.36	11.37	11.37
		0.31	*0.37*	*0.37*	*0.35*	*0.40*	*0.33*	*0.31*	*0.33*	*0.31*	*0.31*
200	0	1.01	1.01	1.20	4.31	2.70	6.97	9.99	11.09	8.70	11.49
		0.08	*0.08*	*0.12*	*0.45*	*0.23*	*0.30*	*0.37*	*0.31*	*0.45*	*0.28*
200	2	3.18	4.16	4.36	4.45	4.39	7.70	5.19	6.56	5.86	11.28
		0.18	*0.27*	*0.28*	*0.29*	*0.27*	*0.33*	*0.36*	*0.34*	*0.40*	*0.32*
200	4	4.94	6.93	6.94	7.65	5.88	8.48	9.47	7.78	10.57	11.01
		0.20	*0.34*	*0.32*	*0.37*	*0.28*	*0.32*	*0.36*	*0.31*	*0.32*	*0.30*
200	6	7.05	10.02	9.79	10.10	9.75	9.32	10.79	9.01	11.05	10.86
		0.24	*0.42*	*0.42*	*0.42*	*0.41*	*0.35*	*0.37*	*0.36*	*0.34*	*0.30*
200	8	9.96	11.46	10.37	11.86	8.81	10.29	11.83	10.03	11.98	11.98
		0.29	*0.34*	*0.33*	*0.32*	*0.37*	*0.32*	*0.31*	*0.31*	*0.32*	*0.32*
200	10	11.08	11.81	11.96	11.30	14.88	11.70	11.13	11.79	11.08	11.08
		0.32	*0.37*	*0.39*	*0.34*	*0.51*	*0.38*	*0.33*	*0.39*	*0.32*	*0.32*
500	0	0.96	1.15	1.50	3.67	2.23	6.44	9.26	9.76	8.67	10.73
		0.09	*0.16*	*0.21*	*0.44*	*0.24*	*0.34*	*0.41*	*0.40*	*0.45*	*0.33*
500	2	3.11	3.71	3.73	3.73	3.86	7.63	4.00	5.16	4.04	11.04
		0.15	*0.21*	*0.21*	*0.21*	*0.21*	*0.29*	*0.23*	*0.25*	*0.24*	*0.28*
500	4	5.23	6.37	6.54	6.62	6.04	8.74	7.71	7.19	7.98	11.43
		0.20	*0.25*	*0.26*	*0.28*	*0.25*	*0.29*	*0.34*	*0.28*	*0.34*	*0.30*
500	6	7.02	9.01	8.87	9.46	7.53	9.32	10.62	8.66	10.98	11.09
		0.22	*0.31*	*0.30*	*0.32*	*0.25*	*0.29*	*0.30*	*0.29*	*0.29*	*0.29*
500	8	9.05	10.92	10.80	10.97	10.34	10.83	11.02	10.58	11.18	11.17
		0.29	*0.31*	*0.31*	*0.31*	*0.35*	*0.31*	*0.32*	*0.30*	*0.32*	*0.32*
500	10	10.96	11.99	12.57	11.39	16.63	12.50	11.50	13.66	10.99	10.96
		0.31	*0.34*	*0.33*	*0.33*	*0.42*	*0.33*	*0.34*	*0.35*	*0.32*	*0.31*

Table C.26: Average Predictive Loss: Normal, $p = 20$, $\rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.04	1.04	1.18	4.91	5.09	12.26	17.36	20.51	15.82	20.95
		0.10	*0.10*	*0.14*	*0.66*	*0.34*	*0.44*	*0.66*	*0.44*	*0.76*	*0.40*
100	4	5.14	9.27	8.35	11.16	8.58	13.97	20.28	14.83	20.68	20.68
		0.21	*0.56*	*0.47*	*0.64*	*0.42*	*0.42*	*0.45*	*0.44*	*0.40*	*0.40*
100	8	9.27	20.42	18.46	21.24	17.77	18.78	21.58	19.26	21.59	21.59
		0.26	*0.51*	*0.47*	*0.46*	*0.44*	*0.47*	*0.44*	*0.47*	*0.44*	*0.44*
100	12	12.58	21.58	22.02	20.98	22.96	19.89	20.96	20.03	20.98	20.98
		0.32	*0.43*	*0.54*	*0.41*	*0.59*	*0.47*	*0.41*	*0.45*	*0.41*	*0.41*
100	16	16.98	21.17	22.77	21.17	28.46	21.30	21.14	21.02	21.17	21.17
		0.36	*0.39*	*0.57*	*0.39*	*0.70*	*0.45*	*0.39*	*0.42*	*0.39*	*0.39*
100	20	21.06	21.06	24.63	21.06	37.82	25.99	21.12	23.11	21.06	21.06
		0.43	*0.43*	*0.61*	*0.43*	*0.64*	*0.52*	*0.43*	*0.47*	*0.43*	*0.43*
200	0	0.94	0.94	1.03	5.27	4.07	12.58	17.66	20.88	16.94	21.23
		0.08	*0.08*	*0.12*	*0.70*	*0.31*	*0.44*	*0.67*	*0.44*	*0.74*	*0.42*
200	4	5.03	6.94	7.05	7.62	7.66	14.56	12.99	13.01	17.01	21.45
		0.21	*0.35*	*0.34*	*0.43*	*0.34*	*0.46*	*0.70*	*0.48*	*0.68*	*0.43*
200	8	9.01	17.58	14.92	19.58	13.37	16.60	21.52	16.31	21.54	21.54
		0.26	*0.61*	*0.55*	*0.54*	*0.54*	*0.43*	*0.41*	*0.44*	*0.42*	*0.42*
200	12	13.52	21.36	20.92	21.31	26.71	19.93	21.29	19.92	21.31	21.31
		0.33	*0.42*	*0.53*	*0.42*	*0.76*	*0.51*	*0.42*	*0.50*	*0.42*	*0.42*
200	16	16.98	21.16	22.65	21.16	39.61	25.12	21.25	24.05	21.16	21.16
		0.38	*0.42*	*0.52*	*0.42*	*0.71*	*0.55*	*0.43*	*0.53*	*0.42*	*0.42*
200	20	20.79	20.79	21.52	20.79	50.93	27.39	22.64	25.61	20.79	20.79
		0.42	*0.42*	*0.45*	*0.42*	*0.86*	*0.56*	*0.40*	*0.55*	*0.42*	*0.42*
500	0	1.09	1.09	1.25	5.15	3.20	12.56	17.81	20.77	16.16	21.11
		0.10	*0.10*	*0.14*	*0.68*	*0.27*	*0.44*	*0.66*	*0.45*	*0.77*	*0.42*
500	4	5.18	6.66	6.72	6.72	7.10	15.01	7.72	11.04	8.12	21.69
		0.21	*0.36*	*0.36*	*0.36*	*0.35*	*0.45*	*0.41*	*0.43*	*0.46*	*0.42*
500	8	9.16	12.62	12.62	12.77	12.40	15.67	17.11	13.50	19.13	20.78
		0.25	*0.50*	*0.48*	*0.50*	*0.51*	*0.41*	*0.55*	*0.40*	*0.51*	*0.39*
500	12	13.30	22.46	21.64	22.30	24.19	20.71	21.47	20.86	21.43	21.43
		0.33	*0.45*	*0.48*	*0.44*	*0.41*	*0.47*	*0.44*	*0.48*	*0.44*	*0.44*
500	16	17.21	21.21	22.53	21.15	37.92	23.05	21.10	24.46	21.15	21.15
		0.38	*0.41*	*0.50*	*0.40*	*0.63*	*0.49*	*0.40*	*0.54*	*0.40*	*0.40*
500	20	20.14	20.14	20.51	20.14	26.12	21.51	20.43	21.99	20.14	20.14
		0.41	*0.41*	*0.41*	*0.41*	*0.70*	*0.44*	*0.43*	*0.47*	*0.41*	*0.41*

Table C.27: Average Predictive Loss: Normal, $p = 50, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
200	0	0.97	0.97	0.97	6.53	7.57	27.93	40.70	49.32	41.36	49.58
		0.09	*0.09*	*0.09*	*1.20*	*0.44*	*0.71*	*1.46*	*0.70*	*1.47*	*0.68*
200	5	6.05	32.73	29.41	47.58	23.78	33.72	50.59	50.55	50.60	50.62
		0.22	*0.47*	*0.37*	*0.82*	*0.56*	*0.70*	*0.63*	*0.63*	*0.63*	*0.63*
200	10	11.24	23.33	19.06	30.36	19.10	35.35	51.99	37.90	52.00	52.00
		0.29	*1.18*	*0.71*	*1.36*	*0.65*	*0.66*	*0.62*	*0.70*	*0.62*	*0.62*
200	15	15.64	20.32	20.22	21.51	23.48	37.37	50.30	39.29	50.79	50.79
		0.32	*0.52*	*0.43*	*0.69*	*0.47*	*0.61*	*0.69*	*0.68*	*0.63*	*0.63*
200	20	21.21	46.99	42.33	48.25	40.96	42.89	51.18	44.59	51.18	51.18
		0.41	*0.85*	*0.77*	*0.75*	*0.71*	*0.66*	*0.61*	*0.68*	*0.61*	*0.61*
200	25	26.06	15.86	16.22	16.48	20.62	36.32	49.19	37.14	50.78	51.09
		0.47	*0.45*	*0.46*	*0.54*	*0.52*	*0.65*	*0.84*	*0.69*	*0.66*	*0.62*
200	30	30.79	50.77	57.69	50.48	60.78	48.20	50.48	48.17	50.48	50.48
		0.49	*0.66*	*0.90*	*0.65*	*0.86*	*0.75*	*0.65*	*0.71*	*0.65*	*0.65*
200	35	36.52	51.21	47.35	51.00	45.46	45.27	51.34	47.05	51.34	51.34
		0.54	*0.70*	*0.80*	*0.64*	*0.75*	*0.66*	*0.60*	*0.65*	*0.60*	*0.60*
200	40	40.23	50.32	63.27	50.32	80.58	56.66	50.33	52.58	50.32	50.32
		0.55	*0.60*	*1.06*	*0.60*	*0.91*	*0.74*	*0.60*	*0.70*	*0.60*	*0.60*
200	45	46.33	51.33	75.08	51.33	81.75	60.97	51.33	55.44	51.33	51.33
		0.61	*0.66*	*1.06*	*0.66*	*0.78*	*0.68*	*0.65*	*0.68*	*0.66*	*0.66*
200	50	50.87	50.87	67.85	50.87	77.68	56.82	50.88	53.36	50.87	50.87
		0.60	*0.60*	*1.17*	*0.60*	*0.92*	*0.75*	*0.60*	*0.66*	*0.60*	*0.60*

Table C.28: Average Predictive Loss: Normal, $p = 10, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.00	1.00	1.15	4.14	2.89	7.22	10.21	10.88	9.90	11.47
		0.09	*0.09*	*0.13*	*0.44*	*0.23*	*0.32*	*0.37*	*0.34*	*0.40*	*0.29*
100	2	3.11	4.61	4.73	5.49	4.93	7.65	7.23	7.32	9.13	11.30
		0.15	*0.27*	*0.26*	*0.35*	*0.26*	*0.32*	*0.42*	*0.33*	*0.41*	*0.30*
100	4	5.00	11.05	9.98	11.01	9.89	9.18	10.96	9.11	10.94	10.92
		0.19	*0.47*	*0.46*	*0.40*	*0.48*	*0.37*	*0.30*	*0.36*	*0.30*	*0.30*
100	6	6.28	3.94	4.12	4.29	4.40	6.75	5.90	6.41	7.59	10.15
		0.23	*0.25*	*0.26*	*0.28*	*0.26*	*0.31*	*0.39*	*0.32*	*0.39*	*0.30*
100	8	9.03	13.73	13.97	12.25	14.75	11.89	11.44	11.86	11.37	11.37
		0.27	*0.36*	*0.41*	*0.34*	*0.39*	*0.35*	*0.31*	*0.35*	*0.30*	*0.30*
100	10	11.15	12.97	16.07	11.56	20.30	14.73	11.49	14.19	11.15	11.15
		0.34	*0.46*	*0.45*	*0.38*	*0.36*	*0.36*	*0.34*	*0.37*	*0.34*	*0.34*
200	0	0.90	0.90	1.09	4.62	2.41	6.61	9.39	10.18	8.15	11.05
		0.08	*0.08*	*0.12*	*0.47*	*0.23*	*0.33*	*0.40*	*0.38*	*0.47*	*0.32*
200	2	2.80	3.90	3.99	4.11	3.95	7.48	4.70	6.16	5.52	10.91
		0.15	*0.26*	*0.27*	*0.29*	*0.25*	*0.34*	*0.35*	*0.34*	*0.41*	*0.33*
200	4	5.12	5.09	5.18	5.18	5.54	8.32	6.07	7.03	6.96	11.38
		0.21	*0.25*	*0.26*	*0.26*	*0.27*	*0.30*	*0.31*	*0.29*	*0.36*	*0.30*
200	6	7.00	12.40	12.19	12.32	12.07	10.55	12.20	10.79	11.95	10.82
		0.25	*0.31*	*0.33*	*0.33*	*0.31*	*0.33*	*0.33*	*0.35*	*0.32*	*0.30*
200	8	9.18	9.68	9.42	10.09	8.69	10.00	10.78	9.69	11.04	11.12
		0.28	*0.36*	*0.34*	*0.35*	*0.33*	*0.32*	*0.33*	*0.31*	*0.32*	*0.32*
200	10	10.51	10.67	11.46	10.53	12.99	11.71	10.64	11.77	10.51	10.51
		0.29	*0.29*	*0.28*	*0.29*	*0.36*	*0.28*	*0.29*	*0.28*	*0.29*	*0.29*
500	0	1.03	1.03	1.19	3.53	2.06	7.01	9.84	10.53	9.41	11.18
		0.09	*0.09*	*0.12*	*0.40*	*0.20*	*0.32*	*0.37*	*0.34*	*0.41*	*0.29*
500	2	3.20	3.78	3.80	3.80	3.91	7.72	4.00	5.13	4.04	11.36
		0.16	*0.21*	*0.21*	*0.21*	*0.21*	*0.31*	*0.23*	*0.28*	*0.23*	*0.29*
500	4	4.88	6.08	6.29	6.35	5.77	8.01	6.92	6.73	7.22	10.60
		0.21	*0.31*	*0.31*	*0.32*	*0.30*	*0.31*	*0.34*	*0.31*	*0.36*	*0.29*
500	6	7.33	11.08	10.80	11.23	11.95	10.38	11.46	10.19	11.77	11.54
		0.25	*0.43*	*0.42*	*0.42*	*0.49*	*0.36*	*0.36*	*0.40*	*0.34*	*0.31*
500	8	8.77	10.88	10.23	10.93	10.98	10.02	10.50	9.93	10.90	10.90
		0.28	*0.33*	*0.32*	*0.31*	*0.53*	*0.32*	*0.30*	*0.36*	*0.31*	*0.31*
500	10	11.03	11.22	12.59	11.13	19.92	13.41	13.44	15.18	11.03	11.03
		0.30	*0.31*	*0.34*	*0.30*	*0.33*	*0.33*	*0.28*	*0.34*	*0.30*	*0.30*

Table C.29: Average Predictive Loss: Normal, $p = 20, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.19	1.19	1.25	5.07	5.25	12.48	17.83	20.92	16.82	21.36
		0.10	*0.10*	*0.11*	*0.67*	*0.33*	*0.43*	*0.66*	*0.45*	*0.74*	*0.41*
100	4	5.30	9.13	9.02	11.62	9.30	14.52	20.30	15.06	21.09	21.22
		0.21	*0.56*	*0.51*	*0.68*	*0.48*	*0.45*	*0.49*	*0.48*	*0.43*	*0.41*
100	8	8.93	21.60	19.87	20.84	19.30	17.92	20.45	18.27	20.47	20.47
		0.26	*0.44*	*0.52*	*0.41*	*0.48*	*0.41*	*0.39*	*0.41*	*0.40*	*0.40*
100	12	12.85	20.75	18.82	20.78	18.04	18.94	20.61	19.41	20.61	20.61
		0.31	*0.46*	*0.45*	*0.42*	*0.41*	*0.39*	*0.39*	*0.38*	*0.39*	*0.39*
100	16	16.67	20.76	26.57	20.65	32.40	23.32	20.61	21.71	20.65	20.65
		0.36	*0.41*	*0.60*	*0.40*	*0.57*	*0.47*	*0.40*	*0.45*	*0.40*	*0.40*
100	20	21.06	9.96	9.64	12.20	10.28	15.15	20.73	15.68	21.06	21.06
		0.42	*0.55*	*0.49*	*0.62*	*0.46*	*0.44*	*0.44*	*0.46*	*0.42*	*0.42*
200	0	0.95	0.95	1.00	5.54	3.43	12.30	17.41	20.74	16.65	21.16
		0.08	*0.08*	*0.09*	*0.72*	*0.29*	*0.46*	*0.69*	*0.46*	*0.76*	*0.43*
200	4	5.13	7.54	7.73	8.30	8.20	15.40	16.80	14.35	20.17	22.62
		0.21	*0.46*	*0.41*	*0.52*	*0.41*	*0.48*	*0.73*	*0.51*	*0.62*	*0.45*
200	8	9.48	11.68	11.91	12.45	12.14	17.38	18.30	16.19	20.50	22.25
		0.27	*0.37*	*0.36*	*0.43*	*0.34*	*0.39*	*0.58*	*0.41*	*0.52*	*0.41*
200	12	12.85	9.53	9.58	10.93	9.95	16.09	17.80	15.38	19.71	21.09
		0.33	*0.40*	*0.37*	*0.49*	*0.34*	*0.42*	*0.59*	*0.43*	*0.51*	*0.42*
200	16	17.15	20.93	20.68	20.93	27.18	20.26	20.90	20.11	20.93	20.93
		0.35	*0.39*	*0.43*	*0.39*	*0.72*	*0.45*	*0.39*	*0.44*	*0.39*	*0.39*
200	20	20.67	24.52	29.39	22.01	34.22	25.88	20.69	25.82	20.67	20.67
		0.38	*0.58*	*0.51*	*0.49*	*0.52*	*0.39*	*0.38*	*0.39*	*0.38*	*0.38*
500	0	1.02	1.02	1.02	3.29	2.76	11.39	16.69	19.82	15.20	20.25
		0.09	*0.09*	*0.09*	*0.52*	*0.24*	*0.41*	*0.62*	*0.41*	*0.72*	*0.37*
500	4	5.09	6.00	6.11	6.11	6.34	13.75	6.92	10.28	7.28	20.70
		0.20	*0.29*	*0.30*	*0.30*	*0.30*	*0.45*	*0.41*	*0.42*	*0.45*	*0.43*
500	8	8.71	18.16	17.59	17.91	18.55	16.87	19.51	15.90	20.60	20.75
		0.25	*0.62*	*0.59*	*0.61*	*0.61*	*0.48*	*0.58*	*0.51*	*0.49*	*0.41*
500	12	13.71	15.15	15.24	15.54	13.75	17.64	20.45	16.33	21.35	21.85
		0.33	*0.56*	*0.54*	*0.55*	*0.51*	*0.47*	*0.51*	*0.49*	*0.46*	*0.40*
500	16	16.88	30.22	32.32	26.36	40.94	25.70	21.39	28.57	21.07	20.90
		0.39	*0.73*	*0.68*	*0.69*	*0.37*	*0.50*	*0.47*	*0.53*	*0.44*	*0.42*
500	20	20.62	20.62	20.87	20.62	39.43	22.91	26.09	23.38	20.62	20.62
		0.40	*0.40*	*0.41*	*0.40*	*1.02*	*0.47*	*0.45*	*0.48*	*0.40*	*0.40*

Table C.30: Average Predictive Loss: Normal, $p = 50, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
200	0	1.06	1.06	1.12	7.94	7.12	27.33	40.69	50.17	41.06	50.41
		0.09	*0.09*	*0.11*	*1.30*	*0.42*	*0.70*	*1.50*	*0.68*	*1.53*	*0.66*
200	5	5.67	8.20	8.60	8.60	12.40	31.15	45.56	31.54	49.84	51.35
		0.22	*0.46*	*0.47*	*0.47*	*0.49*	*0.70*	*1.23*	*0.81*	*0.86*	*0.63*
200	10	10.79	21.79	17.67	30.61	18.02	33.98	50.95	36.27	50.95	50.95
		0.29	*1.16*	*0.75*	*1.39*	*0.72*	*0.65*	*0.60*	*0.68*	*0.60*	*0.60*
200	15	15.66	29.95	29.08	32.13	27.92	35.97	51.35	37.91	51.35	51.35
		0.34	*0.65*	*0.61*	*0.93*	*0.64*	*0.74*	*0.67*	*0.77*	*0.67*	*0.67*
200	20	20.82	43.53	36.13	47.25	35.20	40.74	51.11	42.94	51.11	51.11
		0.38	*0.96*	*0.81*	*0.86*	*0.78*	*0.70*	*0.61*	*0.70*	*0.61*	*0.61*
200	25	26.03	33.96	33.28	34.78	32.58	38.61	51.09	39.97	51.12	51.12
		0.46	*0.60*	*0.56*	*0.70*	*0.55*	*0.63*	*0.62*	*0.65*	*0.62*	*0.62*
200	30	31.28	51.84	48.81	51.29	48.57	46.93	51.23	47.49	51.23	51.23
		0.51	*0.70*	*0.74*	*0.65*	*0.75*	*0.70*	*0.66*	*0.69*	*0.66*	*0.66*
200	35	35.33	53.77	59.75	51.43	57.87	52.31	50.64	51.52	50.65	50.65
		0.53	*0.75*	*0.80*	*0.64*	*0.78*	*0.65*	*0.63*	*0.63*	*0.63*	*0.63*
200	40	40.42	50.77	74.68	50.52	78.15	56.84	50.51	53.73	50.52	50.52
		0.54	*0.63*	*1.04*	*0.61*	*0.80*	*0.71*	*0.61*	*0.69*	*0.61*	*0.61*
200	45	45.94	50.87	66.05	50.87	71.15	56.17	50.85	53.26	50.87	50.87
		0.60	*0.65*	*0.85*	*0.65*	*0.80*	*0.67*	*0.65*	*0.69*	*0.65*	*0.65*
200	50	50.74	50.74	54.88	50.74	64.13	50.58	50.73	50.25	50.74	50.74
		0.70	*0.70*	*0.97*	*0.70*	*1.11*	*0.74*	*0.70*	*0.71*	*0.70*	*0.70*

Table C.31: Percent of Hits: Normal, $p = 10, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.00	1.00	0.98	0.85	0.73	0.18	0.00	0.06	0.30	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.02*	*0.00*	*0.02*	*0.03*	*0.00*
100	2	1.00	0.89	0.86	0.84	0.80	0.28	0.69	0.38	0.46	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
100	4	1.00	0.22	0.28	0.17	0.29	0.24	0.01	0.25	0.00	0.00
		0.00	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*
100	6	1.00	0.00	0.03	0.00	0.01	0.05	0.00	0.05	0.00	0.00
		0.00	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*
100	8	1.00	0.00	0.11	0.00	0.04	0.12	0.00	0.13	0.00	0.00
		0.00	*0.00*	*0.02*	*0.00*	*0.01*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
100	10	1.00	0.62	0.02	0.87	0.00	0.00	0.00	0.01	1.00	1.00
		0.00	*0.03*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*
200	0	1.00	1.00	0.98	0.81	0.78	0.11	0.00	0.02	0.41	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.02*	*0.00*	*0.01*	*0.03*	*0.00*
200	2	1.00	0.89	0.86	0.86	0.84	0.26	0.82	0.50	0.77	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
200	4	1.00	0.76	0.73	0.70	0.88	0.38	0.44	0.54	0.19	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
200	6	1.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*
200	8	1.00	0.00	0.02	0.00	0.00	0.04	0.00	0.02	0.00	0.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*
200	10	1.00	0.73	0.03	0.91	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.03*	*0.01*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	0.99	0.96	0.85	0.87	0.24	0.00	0.12	0.34	0.00
		0.00	*0.01*	*0.01*	*0.02*	*0.02*	*0.03*	*0.00*	*0.02*	*0.03*	*0.00*
500	2	1.00	0.93	0.93	0.93	0.91	0.25	0.90	0.69	0.90	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.02*	*0.00*
500	4	1.00	0.85	0.82	0.82	0.90	0.34	0.70	0.69	0.68	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
500	6	1.00	0.71	0.71	0.65	0.94	0.52	0.28	0.72	0.13	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
500	8	1.00	0.02	0.06	0.00	0.00	0.10	0.00	0.04	0.00	0.00
		0.00	*0.01*	*0.02*	*0.00*	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*	*0.00*
500	10	1.00	0.65	0.17	0.84	0.00	0.06	0.00	0.01	0.99	1.00
		0.00	*0.03*	*0.02*	*0.02*	*0.00*	*0.02*	*0.00*	*0.01*	*0.01*	*0.00*

Table C.32: Percent of Hits: Normal, $p = 20, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.00	1.00	0.99	0.88	0.52	0.06	0.00	0.00	0.36	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.02*	*0.00*	*0.00*	*0.03*	*0.00*
100	4	1.00	0.71	0.71	0.64	0.61	0.06	0.06	0.08	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.01*	*0.02*	*0.00*	*0.00*
100	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	20	1.00	1.00	0.02	1.00	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	1.00	0.86	0.63	0.04	0.00	0.00	0.30	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*
200	4	1.00	0.80	0.78	0.76	0.68	0.07	0.56	0.17	0.36	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*
200	8	1.00	0.40	0.44	0.25	0.62	0.11	0.00	0.18	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
200	12	1.00	0.00	0.14	0.00	0.12	0.17	0.00	0.16	0.00	0.00
		0.00	*0.00*	*0.02*	*0.00*	*0.02*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
200	16	1.00	0.00	0.04	0.00	0.00	0.04	0.00	0.06	0.00	0.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.02*	*0.00*	*0.00*
200	20	1.00	1.00	0.48	1.00	0.00	0.06	0.00	0.11	1.00	1.00
		0.00	*0.00*	*0.03*	*0.00*	*0.00*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
500	0	1.00	1.00	0.99	0.87	0.76	0.04	0.00	0.00	0.35	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*
500	4	1.00	0.87	0.86	0.86	0.80	0.03	0.76	0.30	0.74	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.03*	*0.01*	*0.03*	*0.03*	*0.03*	*0.00*
500	8	1.00	0.68	0.67	0.67	0.73	0.16	0.42	0.41	0.26	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*
500	12	1.00	0.06	0.08	0.02	0.02	0.10	0.00	0.10	0.00	0.00
		0.00	*0.01*	*0.02*	*0.01*	*0.01*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
500	16	1.00	0.00	0.11	0.00	0.00	0.14	0.00	0.14	0.00	0.00
		0.00	*0.00*	*0.02*	*0.00*	*0.00*	*0.02*	*0.00*	*0.02*	*0.00*	*0.00*
500	20	1.00	1.00	0.17	1.00	0.00	0.01	0.00	0.01	1.00	1.00
		0.00	*0.00*	*0.02*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*

Table C.33: Percent of Hits: Normal, $p = 50, \rho = 0$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
200	0	1.00	1.00	1.00	0.92	0.38	0.00	0.00	0.00	0.22	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*
200	5	1.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	10	1.00	0.47	0.43	0.35	0.36	0.00	0.00	0.00	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	15	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	50	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Table C.34: Percent of Hits: Normal, $p = 10, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.00	1.00	0.99	0.82	0.74	0.19	0.00	0.04	0.24	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.03*	*0.02*	*0.00*	*0.01*	*0.03*	*0.00*
100	2	1.00	0.83	0.80	0.78	0.76	0.30	0.65	0.42	0.44	0.00
		0.00	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
100	4	1.00	0.38	0.45	0.29	0.52	0.33	0.04	0.34	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.01*	*0.03*	*0.00*	*0.00*
100	6	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	10	1.00	0.86	0.05	0.96	0.00	0.01	0.00	0.01	1.00	1.00
		0.00	*0.02*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*
200	0	1.00	1.00	0.98	0.79	0.81	0.25	0.00	0.09	0.46	0.00
		0.00	*0.00*	*0.01*	*0.03*	*0.02*	*0.03*	*0.00*	*0.02*	*0.03*	*0.00*
200	2	1.00	0.89	0.88	0.88	0.87	0.28	0.83	0.56	0.78	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
200	4	1.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*
200	6	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	8	1.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00
		0.00	*0.00*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*
200	10	1.00	0.93	0.12	0.99	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.02*	*0.02*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	1.00	0.99	0.86	0.87	0.20	0.00	0.08	0.28	0.00
		0.00	*0.00*	*0.01*	*0.02*	*0.02*	*0.03*	*0.00*	*0.02*	*0.03*	*0.00*
500	2	1.00	0.94	0.93	0.93	0.92	0.26	0.91	0.74	0.90	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.02*	*0.00*
500	4	1.00	0.86	0.82	0.82	0.91	0.43	0.75	0.71	0.73	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
500	6	1.00	0.52	0.52	0.48	0.54	0.42	0.25	0.58	0.06	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.00*
500	8	1.00	0.28	0.63	0.10	0.86	0.69	0.00	0.81	0.00	0.00
		0.00	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.00*	*0.02*	*0.00*	*0.00*
500	10	1.00	0.95	0.46	0.98	0.01	0.26	0.00	0.12	1.00	1.00
		0.00	*0.01*	*0.03*	*0.01*	*0.01*	*0.03*	*0.00*	*0.02*	*0.00*	*0.00*

Table C.35: Percent of Hits: Normal, $p = 20, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
100	0	1.00	1.00	1.00	0.88	0.53	0.04	0.00	0.00	0.31	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*
100	4	1.00	0.74	0.70	0.64	0.61	0.07	0.10	0.12	0.02	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.02*	*0.02*	*0.01*	*0.00*
100	8	1.00	0.00	0.04	0.00	0.04	0.05	0.00	0.03	0.00	0.00
		0.00	*0.00*	*0.01*	*0.00*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*
100	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	20	1.00	0.04	0.00	0.20	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.01*	*0.00*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	1.00	0.85	0.71	0.04	0.00	0.00	0.32	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*
200	4	1.00	0.80	0.75	0.74	0.67	0.07	0.41	0.16	0.21	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*
200	8	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.73	0.00	0.90	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.03*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	1.00	1.00	0.93	0.80	0.06	0.00	0.00	0.36	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.01*	*0.00*	*0.00*	*0.03*	*0.00*
500	4	1.00	0.91	0.90	0.90	0.86	0.12	0.85	0.40	0.84	0.00
		0.00	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.00*
500	8	1.00	0.30	0.31	0.30	0.29	0.11	0.24	0.27	0.12	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.02*	*0.00*
500	12	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	16	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	20	1.00	1.00	0.86	1.00	0.04	0.54	0.00	0.49	1.00	1.00
		0.00	*0.00*	*0.02*	*0.00*	*0.01*	*0.03*	*0.00*	*0.03*	*0.00*	*0.00*

Table C.36: Percent of Hits: Normal, $p = 50, \rho = 0.5$

n	q	gold	cfb	fb.r	fb	bic	aic	cml	mcml	hcml	ns
200	0	1.00	1.00	1.00	0.90	0.42	0.00	0.00	0.00	0.25	0.00
		0.00	*0.00*	*0.00*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*	*0.03*	*0.00*
200	5	1.00	0.80	0.76	0.76	0.38	0.00	0.14	0.00	0.05	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.02*	*0.00*	*0.01*	*0.00*
200	10	1.00	0.56	0.58	0.43	0.48	0.00	0.00	0.00	0.00	0.00
		0.00	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	15	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	20	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	35	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	50	1.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

C.4 Results Under Forward Procedure

Table C.37: Average Predictive Loss: Poisson under FP, $p = 10, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
100	0	0.50	0.55	0.55	0.58	1.35	3.04	2.16	4.82	5.06
		0.04	*0.06*	*0.06*	*0.06*	*0.11*	*0.14*	*0.09*	*0.16*	*0.14*
100	2	2.14	2.14	2.14	2.22	2.44	4.13	2.19	2.49	6.35
		0.16	*0.16*	*0.16*	*0.16*	*0.16*	*0.21*	*0.16*	*0.17*	*0.24*
100	4	2.85	2.89	2.89	3.07	2.99	3.89	3.06	3.12	5.42
		0.17	*0.18*	*0.18*	*0.19*	*0.18*	*0.20*	*0.19*	*0.18*	*0.23*
100	6	12.50	12.20	12.35	12.18	12.07	11.54	12.32	11.36	13.11
		0.55	*0.54*	*0.55*	*0.52*	*0.54*	*0.52*	*0.53*	*0.51*	*0.55*
100	8	7.17	6.92	6.91	6.80	6.89	6.50	6.71	6.74	6.86
		0.30	*0.28*	*0.28*	*0.27*	*0.28*	*0.26*	*0.27*	*0.27*	*0.26*
100	10	6.88	6.91	6.88	6.88	7.34	6.94	13.10	7.11	6.88
		0.32	*0.32*	*0.32*	*0.32*	*0.36*	*0.32*	*0.42*	*0.36*	*0.32*
200	0	1.65	1.75	1.75	2.45	4.15	11.80	8.25	18.46	19.36
		0.15	*0.18*	*0.18*	*0.30*	*0.38*	*0.59*	*0.33*	*0.65*	*0.60*
200	2	1.52	1.50	1.50	1.55	1.87	2.88	1.55	1.83	4.16
		0.10	*0.10*	*0.10*	*0.11*	*0.11*	*0.13*	*0.11*	*0.11*	*0.14*
200	4	4.72	4.75	4.75	5.03	4.89	6.97	5.03	4.99	9.33
		0.24	*0.24*	*0.24*	*0.26*	*0.24*	*0.31*	*0.26*	*0.25*	*0.34*
200	6	5.35	5.26	5.27	5.13	5.29	5.03	5.15	5.09	5.81
		0.21	*0.21*	*0.21*	*0.21*	*0.21*	*0.22*	*0.21*	*0.21*	*0.21*
200	8	5.80	5.83	5.84	6.09	5.58	6.61	6.15	5.77	7.48
		0.23	*0.23*	*0.23*	*0.23*	*0.22*	*0.23*	*0.23*	*0.23*	*0.23*
200	10	6.29	6.31	6.29	6.29	6.89	6.34	18.34	6.49	6.29
		0.28	*0.28*	*0.28*	*0.28*	*0.35*	*0.28*	*0.34*	*0.30*	*0.28*
500	0	0.88	0.95	0.95	1.14	1.76	6.43	4.71	9.91	10.50
		0.09	*0.11*	*0.11*	*0.15*	*0.19*	*0.32*	*0.24*	*0.35*	*0.31*
500	2	2.03	2.02	2.02	2.05	2.20	3.51	2.05	2.20	4.92
		0.15	*0.15*	*0.15*	*0.15*	*0.16*	*0.18*	*0.15*	*0.16*	*0.18*
500	4	3.94	3.95	3.95	4.06	3.96	5.39	4.06	4.11	6.74
		0.23	*0.23*	*0.23*	*0.23*	*0.23*	*0.25*	*0.23*	*0.23*	*0.25*
500	6	3.04	3.07	3.07	3.20	3.02	3.65	3.21	3.04	4.27
		0.11	*0.12*	*0.12*	*0.12*	*0.11*	*0.12*	*0.12*	*0.11*	*0.12*
500	8	6.27	6.27	6.30	6.47	5.88	6.42	6.41	5.93	6.93
		0.22	*0.22*	*0.22*	*0.22*	*0.20*	*0.22*	*0.22*	*0.20*	*0.23*
500	10	3.24	3.24	3.24	3.20	3.26	3.21	3.18	3.25	3.12
		0.11	*0.11*	*0.11*	*0.12*	*0.12*	*0.11*	*0.11*	*0.11*	*0.12*

Table C.38: Average Predictive Loss: Poisson under FP, $p = 20, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
100	0	1.87	2.12	2.12	3.40	12.14	32.10	13.29	60.39	61.41
		0.17	*0.24*	*0.24*	*0.50*	*0.95*	*1.48*	*0.49*	*2.12*	*2.07*
100	4	7.93	8.20	8.20	8.87	10.47	19.54	8.79	13.08	44.07
		0.45	*0.48*	*0.48*	*0.50*	*0.57*	*1.03*	*0.50*	*0.73*	*1.74*
100	8	4.37	4.31	4.31	4.30	4.24	5.37	4.31	4.42	7.51
		0.22	*0.21*	*0.21*	*0.18*	*0.19*	*0.20*	*0.18*	*0.18*	*0.22*
100	12	24.06	23.59	23.59	22.65	21.55	23.32	22.86	21.73	28.80
		0.82	*0.80*	*0.80*	*0.74*	*0.71*	*0.78*	*0.75*	*0.71*	*1.04*
100	16	8.71	8.75	8.79	9.20	8.34	8.80	9.19	8.46	9.28
		0.25	*0.26*	*0.26*	*0.27*	*0.24*	*0.24*	*0.27*	*0.24*	*0.28*
100	20	14.78	15.33	14.52	13.56	16.68	15.87	14.00	16.42	13.46
		0.53	*0.53*	*0.53*	*0.50*	*0.50*	*0.51*	*0.50*	*0.52*	*0.50*
200	0	0.63	0.66	0.66	0.92	3.08	8.95	4.36	14.96	15.22
		0.05	*0.06*	*0.06*	*0.12*	*0.22*	*0.29*	*0.14*	*0.32*	*0.30*
200	4	2.77	2.78	2.78	2.93	3.28	5.38	2.91	3.46	8.42
		0.16	*0.16*	*0.16*	*0.17*	*0.19*	*0.23*	*0.17*	*0.19*	*0.25*
200	8	5.31	5.34	5.34	5.58	6.22	8.68	5.52	6.51	11.55
		0.17	*0.17*	*0.17*	*0.18*	*0.19*	*0.23*	*0.17*	*0.19*	*0.28*
200	12	13.70	14.02	14.02	15.62	12.91	16.37	15.78	13.67	20.28
		0.55	*0.56*	*0.56*	*0.61*	*0.53*	*0.59*	*0.61*	*0.54*	*0.69*
200	16	6.03	5.96	6.11	6.35	5.47	5.83	6.33	5.61	6.38
		0.17	*0.17*	*0.17*	*0.17*	*0.17*	*0.16*	*0.17*	*0.17*	*0.17*
200	20	23.30	23.03	23.03	21.11	22.51	17.11	21.09	21.12	11.50
		0.34	*0.34*	*0.34*	*0.38*	*0.35*	*0.35*	*0.39*	*0.35*	*0.31*
500	0	1.39	1.68	1.68	1.79	3.92	15.70	7.58	25.94	26.56
		0.13	*0.20*	*0.20*	*0.21*	*0.35*	*0.58*	*0.24*	*0.63*	*0.59*
500	4	2.56	2.60	2.60	2.69	2.89	5.70	2.65	3.07	8.40
		0.14	*0.14*	*0.14*	*0.14*	*0.15*	*0.19*	*0.14*	*0.15*	*0.19*
500	8	6.16	6.25	6.25	6.54	6.23	9.27	6.54	6.78	12.52
		0.25	*0.26*	*0.26*	*0.27*	*0.25*	*0.32*	*0.27*	*0.28*	*0.35*
500	12	7.01	7.09	7.09	7.43	6.78	8.46	7.45	7.05	10.06
		0.22	*0.22*	*0.22*	*0.23*	*0.21*	*0.24*	*0.23*	*0.22*	*0.25*
500	16	11.98	11.75	11.77	10.51	14.69	10.06	10.49	12.82	9.90
		0.31	*0.31*	*0.31*	*0.24*	*0.31*	*0.23*	*0.24*	*0.29*	*0.20*
500	20	8.93	9.33	8.91	8.86	13.19	9.65	10.69	11.82	8.85
		0.21	*0.22*	*0.21*	*0.21*	*0.22*	*0.23*	*0.22*	*0.24*	*0.21*

Table C.39: Average Predictive Loss: Poisson under FP, $p = 50, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
200	0	1.71	2.00	2.00	2.36	13.05	51.18	11.99	105.53	105.84
		0.15	*0.25*	*0.25*	*0.32*	*0.79*	*1.67*	*0.34*	*2.44*	*2.41*
200	5	3.00	3.04	3.04	3.27	5.03	12.74	3.23	6.11	28.67
		0.16	*0.16*	*0.16*	*0.17*	*0.22*	*0.37*	*0.17*	*0.26*	*0.52*
200	10	6.63	6.71	6.71	7.13	8.15	18.17	7.10	10.73	38.47
		0.28	*0.28*	*0.28*	*0.29*	*0.31*	*0.56*	*0.29*	*0.39*	*0.94*
200	15	24.45	23.71	23.71	21.11	18.27	21.30	21.40	17.13	35.17
		1.06	*1.03*	*1.03*	*0.91*	*0.74*	*0.60*	*0.91*	*0.58*	*0.66*
200	20	37.56	37.46	37.46	36.91	36.03	39.15	37.04	36.75	54.16
		0.65	*0.65*	*0.65*	*0.63*	*0.61*	*0.90*	*0.64*	*0.65*	*1.16*
200	25	27.69	27.34	27.34	26.38	24.91	27.44	26.34	24.19	38.78
		0.65	*0.65*	*0.65*	*0.65*	*0.60*	*0.60*	*0.64*	*0.57*	*0.75*
200	30	58.78	58.22	58.22	52.88	63.01	48.63	53.05	51.54	58.07
		1.72	*1.72*	*1.72*	*1.51*	*1.62*	*1.28*	*1.54*	*1.47*	*1.32*
200	35	71.94	71.03	71.03	66.77	73.22	61.75	66.88	64.87	64.66
		1.45	*1.43*	*1.43*	*1.49*	*1.36*	*1.43*	*1.49*	*1.46*	*1.53*
200	40	44.35	43.54	43.42	34.00	60.77	39.15	34.06	45.43	33.61
		1.09	*1.07*	*1.07*	*0.70*	*1.04*	*0.76*	*0.71*	*0.88*	*0.68*
200	45	34.82	34.74	34.74	35.58	35.49	35.16	35.59	33.72	44.81
		0.89	*0.90*	*0.90*	*0.87*	*0.89*	*0.77*	*0.86*	*0.79*	*0.97*
200	50	18.52	18.46	18.37	17.81	25.33	18.58	17.81	19.79	17.72
		0.37	*0.35*	*0.35*	*0.32*	*0.62*	*0.35*	*0.33*	*0.38*	*0.33*

Table C.40: Percent of Hits: Poisson under FP, $p = 10, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
100	0	1.00	0.99	0.99	0.99	0.74	0.19	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.02*	*0.00*	*0.00*	*0.00*
100	2	0.96	0.96	0.96	0.94	0.85	0.29	0.94	0.84	0.00
		0.01	*0.01*	*0.01*	*0.02*	*0.02*	*0.03*	*0.01*	*0.02*	*0.00*
100	4	0.92	0.92	0.92	0.86	0.86	0.44	0.86	0.82	0.00
		0.02	*0.02*	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
100	6	0.29	0.33	0.32	0.33	0.32	0.36	0.31	0.42	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
100	8	0.02	0.03	0.03	0.04	0.00	0.04	0.00	0.01	0.00
		0.01	*0.01*	*0.01*	*0.01*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*
100	10	1.00	0.99	1.00	1.00	0.90	0.98	0.00	0.95	1.00
		0.00	*0.01*	*0.00*	*0.00*	*0.02*	*0.01*	*0.00*	*0.01*	*0.00*
200	0	1.00	1.00	1.00	0.96	0.81	0.17	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.01*	*0.02*	*0.02*	*0.00*	*0.00*	*0.00*
200	2	0.95	0.96	0.96	0.94	0.78	0.26	0.94	0.80	0.00
		0.01	*0.01*	*0.01*	*0.02*	*0.03*	*0.03*	*0.02*	*0.03*	*0.00*
200	4	0.95	0.94	0.94	0.89	0.91	0.39	0.89	0.88	0.00
		0.01	*0.01*	*0.01*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.00*
200	6	0.47	0.50	0.49	0.52	0.48	0.48	0.51	0.56	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*
200	8	0.01	0.01	0.01	0.02	0.01	0.04	0.02	0.01	0.00
		0.01	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.00*
200	10	1.00	1.00	1.00	1.00	0.94	0.99	0.00	0.98	1.00
		0.00	*0.00*	*0.00*	*0.00*	*0.02*	*0.01*	*0.00*	*0.01*	*0.00*
500	0	1.00	0.99	0.99	0.98	0.89	0.20	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.02*	*0.03*	*0.00*	*0.00*	*0.00*
500	2	0.98	0.98	0.98	0.97	0.90	0.26	0.97	0.90	0.00
		0.01	*0.01*	*0.01*	*0.01*	*0.02*	*0.03*	*0.01*	*0.02*	*0.00*
500	4	0.96	0.96	0.96	0.93	0.95	0.37	0.93	0.90	0.00
		0.01	*0.01*	*0.01*	*0.02*	*0.01*	*0.03*	*0.02*	*0.02*	*0.00*
500	6	0.95	0.94	0.94	0.88	0.96	0.50	0.87	0.95	0.00
		0.01	*0.02*	*0.02*	*0.02*	*0.01*	*0.03*	*0.02*	*0.01*	*0.00*
500	8	0.88	0.87	0.86	0.77	1.00	0.77	0.68	0.98	0.00
		0.02	*0.02*	*0.02*	*0.03*	*0.00*	*0.03*	*0.03*	*0.01*	*0.00*
500	10	0.14	0.11	0.16	0.31	0.02	0.16	0.00	0.03	1.00
		0.02	*0.02*	*0.02*	*0.03*	*0.01*	*0.02*	*0.00*	*0.01*	*0.00*

Table C.41: Percent of Hits: Poisson under FP, $p = 20, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
100	0	1.00	0.99	0.99	0.96	0.53	0.03	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
100	4	0.88	0.86	0.86	0.79	0.62	0.10	0.80	0.42	0.00
		0.02	*0.02*	*0.02*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
100	8	0.66	0.66	0.66	0.58	0.61	0.17	0.58	0.52	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.00*
100	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
100	16	0.01	0.02	0.01	0.00	0.00	0.02	0.00	0.01	0.00
		0.01	*0.01*	*0.01*	*0.00*	*0.00*	*0.01*	*0.00*	*0.01*	*0.00*
100	20	0.68	0.11	0.76	0.98	0.00	0.01	0.00	0.00	1.00
		0.03	*0.02*	*0.03*	*0.01*	*0.00*	*0.01*	*0.00*	*0.00*	*0.00*
200	0	1.00	1.00	1.00	0.97	0.60	0.03	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.01*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
200	4	0.94	0.94	0.94	0.88	0.72	0.09	0.90	0.64	0.00
		0.01	*0.02*	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.03*	*0.00*
200	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	12	0.76	0.72	0.72	0.49	0.87	0.28	0.46	0.73	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*
200	16	0.45	0.46	0.38	0.11	0.87	0.49	0.06	0.76	0.00
		0.03	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.02*	*0.03*	*0.00*
200	20	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	1.00
		0.00	*0.00*	*0.00*	*0.02*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
500	0	1.00	0.98	0.98	0.98	0.78	0.04	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*
500	4	0.97	0.96	0.96	0.93	0.86	0.07	0.94	0.78	0.00
		0.01	*0.01*	*0.01*	*0.02*	*0.02*	*0.02*	*0.01*	*0.03*	*0.00*
500	8	0.88	0.87	0.87	0.78	0.87	0.13	0.78	0.72	0.00
		0.02	*0.02*	*0.02*	*0.03*	*0.02*	*0.02*	*0.03*	*0.03*	*0.00*
500	12	0.80	0.77	0.77	0.65	0.89	0.22	0.65	0.78	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.02*	*0.03*	*0.03*	*0.03*	*0.00*
500	16	0.16	0.18	0.17	0.20	0.03	0.27	0.19	0.07	0.00
		0.02	*0.02*	*0.02*	*0.03*	*0.01*	*0.03*	*0.02*	*0.02*	*0.00*
500	20	0.96	0.65	0.98	1.00	0.06	0.51	0.00	0.15	1.00
		0.01	*0.03*	*0.01*	*0.00*	*0.01*	*0.03*	*0.00*	*0.02*	*0.00*

Table C.42: Percent of Hits: Poisson under FP, $p = 50, \rho = 0$

Forward Procedure

n	q	cfb	fb.r	fb	hcml	bic	aic	cml	mcml	ns
200	0	1.00	0.99	0.99	0.98	0.39	0.00	0.00	0.00	0.00
		0.00	*0.01*	*0.01*	*0.01*	*0.03*	*0.00*	*0.00*	*0.00*	*0.00*
200	5	0.91	0.90	0.90	0.84	0.40	0.01	0.85	0.29	0.00
		0.02	*0.02*	*0.02*	*0.02*	*0.03*	*0.01*	*0.02*	*0.03*	*0.00*
200	10	0.80	0.78	0.78	0.69	0.47	0.00	0.69	0.15	0.00
		0.03	*0.03*	*0.03*	*0.03*	*0.03*	*0.00*	*0.03*	*0.02*	*0.00*
200	15	0.16	0.16	0.16	0.17	0.20	0.01	0.17	0.11	0.00
		0.02	*0.02*	*0.02*	*0.02*	*0.03*	*0.01*	*0.02*	*0.02*	*0.00*
200	20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
200	50	0.49	0.00	0.58	0.97	0.00	0.00	0.00	0.00	1.00
		0.03	*0.00*	*0.03*	*0.01*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer, second edition.

Bleistein, N. & Handelsman, R. A. (1975). *Asymptotic Expansions of Integrals.* New York: Holt, Rinehart and Winston.

Clyde, M. & DeSimone-Sasinowska, H. (1998). *Accounting for Model Uncertainty in Poisson Regression Models: Does Particulate Matter Particularly Matter?* Technical Report 97-06, Institute of Statistics and Decision Sciences, Duke University.

Clyde, M. A. (1999). Bayesian model averaging and model search strategies. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics*, 6 (pp. 157–185). Oxford: University Press.

Dellaportas, P. & Forster, J. J. (1999). Markov chain monte carlo model determination for hierarchical and graphical log-linear models. *Biometrika*, 86, 615–633.

Dellaportas, P., Forster, J. J., & Ntzoufras, I. (2000). Bayesian variable selection using the gibbs sampler. In D. K. Dey, S. Ghosh, & B. Mallick (Eds.), *Generalised linear models: A Bayesian perspective* (pp. 271–286). New York: Marcel Dekker.

Dellaportas, P., Forster, J. J., & Ntzoufras, I. (2002). On bayesian model and variable selection using mcmc. *Statistics and Computing*, 12, 27–36.

Draper, N. R. & Smith, H. (1981). *Applied Regression Analysis*. New York: Wiley, second edition.

Efstathiou, M., Gutierrez-pena, E., & Smith, A. E. M. (1998). Laplace approximations for natural exponential families with cuts. *Scandinavian Journal of Statistics*, 25, 77–92.

Fahrmeir, L. & Tutz, G. (1994). *Multivariate Statistical Modeling Based on Generalized Linear Models*. New York: Springer.

Foster, D. P. & George, E. I. (1994). The risk inflation criterion for multiple regression. *Annals of Statistics*, 22, 1947–1975.

George, E. I. (2000). The variable selection problem (vignette).

George, E. I. & Foster, D. P. (2000). Calibration and empirical bayes variable selection. *Biometrika*, 87, 731–747.

George, E. I. & McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88, 881–890.

George, E. I. & McCulloch, R. E. (1996). Stochastic search variable selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 203–214). London: Chapman and Hall.

George, E. I. & McCulloch, R. E. (1997). Approaches for bayesian variable selection. *Statist. Sinica.*, 7, 339–373.

George, E. I., McCulloch, R. E., & Tsay, R. (1994). Two approaches to bayesian model selection with applications. In D. A. Berry, K. M. Chaloner, & J. F. Geweke (Eds.), *Bayesian Statistics and Econometrics: Essays in Honor of A. Zellner*. New York.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32, 1–51.

Hoeting, J., Madigan, D., & Raftery, A. E. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14, 382–417. with discussion.

Jorgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Serial B*, 49, 127–162.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.

Kass, R. E., Tierney, L., & Kadane, J. B. (1990). The validity of posterior expansions based on laplace's method. In S. Geisser, J. S. Hodges, S. J. Press, & A. Zellner (Eds.), *Bayesian and Likelihood Methods in Statistics and Econometrics* (pp. 473–483). Amsterdam: Elsevier Science.

Kass, R. E. & Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90, 938–934.

Leamer, E. (1978). *Specication Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Lindsey, J. K. (1997). *Applying Generalized Linear Models*. New York: Springer.

Madigan, D. M. & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *J.Amer. Statist. Assoc.*, 89, 1535–1546.

Madigan, D. M., Raftery, A. E., Volinsky, C., & Hoeting, J. (1996). Bayesian model averaging. In P. Chan, S. Stolofo, & D. Wolpert (Eds.), *Integrating Multiple Learned Models (IMLM-96)* (pp. 77–83).

Mallows, C. L. (1973). Some comments on c_p. *Technometrics*, 15, 661–676.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall, second edition.

Miller, A. J. (1990). *Subset Selection in Regression*. New York: Chapman-Hall.

Mitchell, T. J. & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression (with discussion). *Journal of the American Statistical Association*, 83, 1023–1036.

Morris, C. (1983a). Parametric empirical bayes confidence sets. In G. E. P. Box, T. Leonard, & C. F. Wu (Eds.), *Scientific Inference, Data Analysis, and Robustness*. New York: Academic Press.

Morris, C. (1983b). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78, 47–65.

Ntzoufras, I., Dellaportas, P., & Forster, J. J. (2002). Bayesian variable and link determination for generalised linear models. *Journal of Statistical Planning and Inference*. to appear.

Raftery, A. (1996). Approximate bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika*, 83, 251–266.

Raftery, A. E., Madigan, D. M., & Hoeting, J. (1997). Model selection and accounting for model uncertainty in linear regression models. *J.Amer.Statist.Assoc.*, 92, 179–191.

Raftery, A. E., Madigan, D. M., & Volinsky, C. (1996). Accounting for model uncertainty in survival analysis improves predictive performance. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics*, 5 (pp. 323–330). Oxford: Univeristy Press.

Raftery, A. E. & Richardson, S. (1993). Model selection for generalized linear models via glib, with application to epidemiology. In D. A. Berry & D. K. Stangl (Eds.), *Bayesian Biostatistics*. New York: Marcel Dekker.

Robbins, H. (1955). An empirical bayes approach to statistics. In *Third Berkeley Symp. Math. Statist. Probab.*, 1955 (pp. 157–164). Berkeley: University of California Press.

Robbins, H. (1964). The empirical bayes approach to statistical decision problems. *Ann. Math. Statist.*, 35, 1–20.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Smith, M. & Kohn, R. (1996). Nonparametric regression using bayesian variable selection. *Journal of Econometrics*, 75, 317–344.

Spiegelhalter, D. J., Dawid, A., Lauritzen, S., & Cowell, R. (1993). Bayesian analysis in expert systems (with discussion). *Statis. Sci.*, 8, 219–283.

Thompson, M. L. (1978). Selection of variables in multiple regression. *International Statistical Review*, 46, 1–19 and 129–146.

Tierney, L. & Kadane, J. B. (1986). Accurate approximation for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82–86.

Wei, B. C. (1997). *Lecture Notes in Statistics: Exponential Family Nonlinear Models*. Singapore: Springer.

Weisberg, S. (1985). *Applied Linear Regression*. New York: Wiley.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In P. K. Goel & A. Zellner (Eds.), *Bayesian Inference and Decision Techniques-Essays in Honor of Bruno de Finetti* (pp. 233–243). Amsterdam: North-Holland.

Zwillinger, D. (1992). *Handbook of Integration.* Boston: Jones and Bartlett Publishers.

Vita

Xinlei Wang was born on November 2, 1975 in Anhui Province , P.R.China, the daughter of Niansheng Wang and Meihua Ju. She graduated from the University of Science and Technology of China in 1997 with a B.S. in Automatic Control. In 1998, she received her M.A. in Statistics from the University of Pittsburgh. In December 2002, she will receive her Ph.D. in Decision Science and Statistics from the University of Texas at Austin.

Permanent address: 2501 Lake Austin Blvd, K201
Austin, Texas 78703

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.